Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Financial Development and Wealth Inequality: A Panel Data Analysis

Author: **Bc. Paul Mainka**

Supervisor: **Prof. Roman Horváth, Ph.D.**

Academic Year: **2020/2021**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently; using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Berlin, January 04, 2021

Signature

Acknowledgments

I would like to thank my supervisor Prof. Roman Horváth, who provided me with valuable advice throughout the work on my thesis. Furthermore, I want to thank the teachers of the Master thesis seminar, Prof. Michal Mejstřík, Dr. Zuzana Havránková and especially Dr. Tomáš Havránek, who stimulated me to work continuously and offered interesting suggestions and advice.

Likewise, my gratitude goes to Prof. Jonathan Powell of the University of Central Florida and Fabio Comelli of the International Monetary Fund for devoting time to personally answer my questions about their data sets, and in particular to Dr. Anthony Shorrocks for going one step further and being so kind as to share his personal data set of the Credit Suisse Wealth Reports with me.

In addition, I would like to express my gratitude to Dr. Georg Herbstritt and Eugene Collins for proofreading my work and refining my writing with me. Last but not least, I am grateful for the moral support of my family and friends.

This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 681228.

Abstract

The understanding of the drivers of wealth inequality is still relatively limited, which is due to the hitherto rather scarce available data on wealth. In the wealth inequality theory, saving is a crucial determinant of wealth inequality, and therefore my thesis emphasizes saving, which I approximate by financial development.

Through a relatively new wealth panel dataset, I dispose of data on the Gini coefficient of wealth for 129 countries over the period 2000 to 2018. I identify the likely most influential variables on wealth inequality from a broad pool of possible explanatory variables by employing lasso for fixed effects and subsequently quantify their effect through fixed effects modelling.

I obtain robust results that globalisation and a business-friendly regulatory environment are associated with higher wealth inequality, while a higher labour force participation rate and stronger control of corruption are linked to lower inequality. Moreover, but slightly less robustly, I find that a greater depth of financial markets is associated with higher wealth inequality.

Thus, I do not find clear empirical support for the prominent role of saving for wealth inequality which it is attributed in theory. Instead, non-financial variables appear to be more relevant.

JEL Classification	C33, E21, G51
Keywords	Wealth inequality, financial development, Gini coefficient

Author’s e-mail	paulmaink@gmail.com
Supervisor’s e-mail	roman.horvath@fsv.cuni.cz

Content

List of Tables

List of Figures

Master's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University

Author:	**Paul Mainka**	Supervisor:	Prof. Roman Horváth Ph.D.
E-mail:	75650712@fsv.cuni.cz	E-mail:	roman.horvath@fsv.cuni.cz
Phone:	+4915123324886	Phone:	222 112 317
Specialization:	MEF	Defense Planned:	September 2020

Proposed Topic:

Financial Development and Wealth Inequality: A Panel Data Analysis

Motivation:

The importance of wealth compared to income has increased as Zucman and Piketty (2014) show that the wealth-income ratios in six of the most developed countries (United States, United Kingdom, France, Germany, Italy and Japan) grew from around 200–300% in 1970 to 400–600% in 2010. Additionally, Davies et al. (2017) find that since the global financial crisis 2007 until 2014 wealth inequality has been rising. During this period, the share of the total wealth of the top decile and the top percentile of the wealthiest people grew in all regions of the world except North America. Yet, considerable differences remain between countries in terms of wealth inequality, whether industrialized or developing. For instance, in Hong Kong, the USA, China or Russia the top decile possesses more than 70% of the wealth while in Slovakia, Belgium or Japan the top decile holds less than 50%.

However, little is known about what causes the differences between countries as there are only few studies on the determinants of wealth inequality. This is due primarily to in the past inadequate availability of data on wealth. Attempts have been made to explain wealth inequality by theoretical models (Piketty and Zucman 2014) or by dynamic quantitative models (Nardi and Fella 2017). There are also empirical studies, which were carried out with a small sample of mostly developed countries with a good availability of data (Alvaredo et al. 2013).

Mares et al. (2018) conduct a broader study including a set of 73 countries using data from the newly obtainable Credit Suisse Global Wealth Report. By utilizing Bayesian Model Averaging as their methodological framework they find that among 37 different economic, financial, political and social determinants only seven are significant and explain about half of the variation in wealth inequality (measured as wealth Gini coefficient). According to their results more finance, globalization and war increase inequality while more efficiency and greater access to finance, redistribution and education decrease it. The fact that three out of seven factors can be attributed to finance and that they affect wealth inequality differently demonstrates finance importance and complexity.

Hypotheses:

1. Hypothesis #1: A higher level of financial development is associated with a lower level of wealth inequality.

2. Hypothesis #2: More efficiency and greater access to finance decrease wealth inequality.
3. Hypothesis #3: More finance (i.e., large financial markets and financial institutions) increase wealth inequality.

Methodology:

In the first part of my study, I will compare the work already conducted on wealth inequality and summarize the most important findings.

The core of my own work will be to replicate the results of Mares et al. (2018) using a different methodology. I will use Panel Data Approach and/or GMM. My panel will cover the period from 2010 to 2019 with annual observations (i.e. 9 time periods) and about 200 countries.

As the dependent variable, I use the Gini coefficient for wealth, which is published every year in October since 2010 in the Databook of the Global Wealth Report of Credit Suisse (the last report I will include will be published in October 2019).

For the independent variables, I focus on those that Mares et al. (2018) estimated to be significant and specifically on the three finance variables. The significant variables are more finance, globalization, war, more efficiency in finance, greater access to finance, redistribution and education. I will use Data for the financial variables from the Global Financial Development Database (GFDD) of the World Bank.

Expected Contribution:

Most explanations for wealth inequality are based on models, as until recently no far-reaching data were available. Empirical studies to date have generally been conducted using a limited number of countries. My study is based on a large data set including a wide range of highly diverse countries and covers the most recent decade. An existing study using this data and the methodology I use is in my knowledge not available.
My study should contribute to a better understanding of the reasons for wealth inequality and the differences in levels of inequality between countries utilizing previously non-existent and up-to-date data. My focus will be to assess the validity of the results of Mares et al. (2018), in particular the impact of the three financial factors, in another methodological framework.

By providing a scientific basis for the social and political debate on inequality, my findings can contribute to factualizing the discussion. Understanding the causes of wealth inequality allows the creation of effective ways to reduce it, if this is politically desired.

Outline:

1. Introduction: Motivation that wealth-inequality increased in the recent years and significant difference in the level of inequality between countries are observable but that understanding of the determinants is still low
2. Literature review: I will present different approaches of understanding wealth-inequality from theoretical models to other empirical studies. A focus will be to present the study of Mares, Hasan and Horvath to which I will refer in the further course of my work.
3. Data: Present the data I use

4. Methodology: I describe the methods I employ, Panel Data Models and GMM
5. Results: I discuss my results including a robustness check and compare them with the results of the study of Mares, Hasan and Horvath
6. Conclusion: I summarize my findings and give an outlook for further research possibilities and practical use

Core Bibliography:

Alvaredo, Facundo; Atkinson, Anthony B.; Piketty, Thomas; Saez, Emmanuel (2013): The Top 1 Percent in International and Historical Perspective. In Journal of Economic Perspectives 27 (3), pp. 3–20. DOI: 10.1257/jep.27.3.3.

Davies, James B.; Lluberas, Rodrigo; Shorrocks, Anthony F. (2017): Estimating the Level and Distribution of Global Wealth, 2000-2014. In Review of Income and Wealth 63 (4), pp. 731–759. DOI: 10.1111/roiw.12318.

Mares, Jan; Horvath, Roman; Hasan, Iftekhar (2018): Finance and Wealth Inequality. IES FSV, Charles University. Prague (IES Working Papers, 35). Available online at http://ies.fsv.cuni.cz.

Nardi, Mariacristina de; Fella, Giulio (2017): Saving and wealth inequality. In Review of Economic Dynamics 26, pp. 280–300. DOI: 10.1016/j.red.2017.06.002.

Piketty, Thomas; Zucman, Gabriel (2014): Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010 *. In The Quarterly Journal of Economics 129 (3), pp. 1255–1310. DOI: 10.1093/qje/qju018.

Global Financial Development Database (GFDD), The World Bank.

Global Wealth Databook 2010 -2018, Credit Suisse

Author **Supervisor**

1. Introduction

In August 2020, amidst the most severe health and economic crisis of the post-World War II era, Jeff Bezos was the first person ever to surpass a net worth of 200 billion US dollars (Toh, 2020). In the same year in which the economic crisis associated with Covid-19 is estimated to push 88 million to 115 million people worldwide into extreme poverty (The World Bank, 2020h), the wealth of the five richest people in the world has soared by a total of more than 250 billion US dollars[1] (Bloomberg L.P., 2020). While these are extreme illustrations, the new Corona virus exposed existing wealth inequality in an unfiltered way.

The Covid 19 crisis appears to have exacerbated wealth inequality and brought the subject into further focus. However, it is an issue that has been in constant discussion within academic circles and the public sphere, particularly since the publication of Piketty's widely circulated book "Capital in the 21st Century" (Piketty and Goldhammer, 2014).

Yet, despite the relevance of the subject, wealth and the fundamentals of how and why it is distributed are not particularly well understood compared to other economic variables. This is primarily caused by the scarcity of comprehensive data on wealth, which is due to the inherent difficulty of estimating the value of a stock variable such as wealth[2] and the reluctance of especially the wealthy segments of the population to disclose the size of their wealth holdings.

The knowledge available to date stems from the existing scientific literature which can be roughly divided into two categories: Descriptive Studies that primarily attempt to compile detailed data on wealth and its distribution, and theoretical studies that construct models and optimise them to produce distributions of wealth that are comparable to those of the scarce real world data. The main insight of the descriptive literature is that, at least for developed countries, wealth inequality has been following an increasing trend since the 1980s (Piketty and Zucman, 2014; Roine and Waldenström, 2015), while in the majority of theoretical studies, saving assumes a key role.

In the majority of theoretical models the underlying concept is that households seek to optimise their consumption over time and achieve this through saving. Wealth inequality then arises as a result of households differing either in the intensity of their motivation to save (Hendricks,

---

[1] Between 01.01.2020 and 31.12.2020. As of 31.12.2020

[2] Since for a stock variable only a hypothetical market value can be computed and alternative measures such as the book value or the tax value do not provide a reliable approximation of the market value nor can serve as a reliable comparable value.

2007), in their ability to save (Bozio et al., 2017) or because of their initial conditions (Huggett et al., 2011).

Due to the prominent structural function of saving in the theory on wealth inequality, I will emphasise saving in my thesis. According to microeconomic theory, finance in the form of financial markets and financial institutions facilitate the ability of individuals and society as a whole to optimise their consumption according to their preferences through saving (Freixas and Rochet, 1997). Hence, finance in theory enables the full realisation of the differences in saving which are causing wealth inequality in the models. Especially on a societal or country level, such as in my study, financial development therefore represents the closest available approximation of saving.

Thereby, diverse effects of financial development on wealth inequality can be expected, since different aspects of financial development tend to disproportionately benefit varying parts of the population. For instance, greater access to finance for people at the lower end of the wealth distribution may reduce wealth inequality, while more investment opportunities in the financial markets may benefit predominantly the wealthy and thus increase inequality (Mares et al., 2018).

Prior to my thesis, only one study had empirically investigated the determinants of wealth inequality across countries (Mares et al., 2018). Thereby, Mares et al. (2018) found confirmation of the importance of the influence of financial development on wealth inequality. Moreover, the authors identify mixed effects of financial development, whereby higher efficiency and access to finance decrease wealth inequality, while greater depth of the financial markets enhance it.

Their study was made possible by the recent provision of a comprehensive dataset on wealth, covering the period from 2000 to 2019 and 172 countries (Davies et al., 2017). In my thesis, I employ the Gini coefficient of wealth as the measure of wealth inequality obtained from the same panel data set. The Gini coefficient is a well-established measure of the distribution of economic variables and enables the comparisons of the distribution of wealth across countries of unequal size and prosperity.

As so far very little established empirical evidence exists on the determinants of wealth inequality, the primary aim of my work is to generate initial empirical insights by identifying relevant explanatory variables and assessing their influence on wealth inequality. Thereby I proceed in two steps.

First, I identify the likely most influential variables from a pool of possible variables using the least absolute shrinkage and selection operator (lasso) for fixed effects. The pool of possible explanatory variables comprises, in addition to various indicators of financial development, several other variables cited in the literature as being important for wealth inequality, such as education or globalisation.

I employ lasso for fixed effects to address my model selection problem instead of the more conservative approach of hypothesis testing for two main reasons: to avoid pre-test biases in hypothesis testing (Giles and Giles, 1993) and because lasso allows my results to be more replicable as it reduces my 'researcher degrees of freedom' and ability to 'p-hack'. Additionally, lasso enables me to consider a larger number of possible explanatory variables than would be feasible through the manual hypothesis testing. Furthermore, lasso selects variables under the constraint of sparsity of the selected model, thus reducing the risk of multicollinearity.

In a second step, I analyse the relationship between the selected variables and wealth inequality using the fixed effects model. For my initial exploration of the influence of the determinants on wealth inequality, the fixed effects model is well suited as its properties, advantages and disadvantages are well understood. Furthermore, of the standard panel data models, fixed effects can be expected to be the efficient and consistent estimator, since I am studying a wide variety of countries and thus the likelihood of the existence of time-invariant omitted variables is high.

In several ways, my paper represents novelty. Firstly, I am merely the second after Mares et al. (2018) to conduct an empirical cross-country analysis of wealth inequality. Thus, while the theoretical models provide guidance on which variables might be relevant, I supply empirical evidence on concrete variables. Furthermore, I am the first to account for the panel structure of the data, particularly the time dimension[3]. Finally, I utilise a considerably larger proportion of the wealth data than Mares et al. (2018) in their study[4].

My paper is structured as follows: first, I present in chapter 2 ‘Literature Review’ the relevant literature in more detail, followed by a discussion of the properties of my wealth data in chapter 3 ‘Data’. Next, I carry out the lasso estimation in chapter 4 ‘Lasso’ and subsequently provide

[3] Mares et al. (2018) average the (wealth) data between the years 2010 and 2016, thus eliminating the time dimension in their work. Consequently, they cannot account for the within variation, the change in wealth inequality over time.
[4] Mares et al. (2018) employ 73 countries and data over the period between 2010 and 2016 in comparison with 129 countries and the period between 2000 and 2018 in my baseline regression.

in chapter 5 ‘Descriptive Analysis’ a detailed descriptive analysis of the independent variables selected by lasso and of my dependent variable the Gini coefficient of wealth. Afterwards, I conduct in chapter 6 ‘Baseline regression’ my fixed effects regression and present the results, along with robustness checks. Subsequently, in the chapter 7 ‘Discussion’, I interpret my results and put them in the scientific context and then as a final step formulate my conclusions in chapter 8 ‘Conclusion’.

2. Literature Review

In this section I examine three different types of literature. The first explores the levels, distribution and historical evolution of wealth. The second examines studies that primarily employ theoretical models to investigate the causes of wealth inequality. Finally, I examine a series of papers that conducted empirical studies on income inequality, a subject closely related to wealth inequality. These papers were of interest because the model and subject of study have parallels with this thesis.

Despite different approaches and definitions, the bulk of the literature investigating the historical development of levels and distribution of wealth comes to similar conclusions. Due to data restrictions, the countries studied are mostly developed countries. The overall picture appears to be that the importance of wealth in the studied countries has decreased from the end of about the second world war until broadly the 1980s and has increased since then.

This is reflected in the movement of wealth to income ratios and the development of wealth inequality. Piketty and Zucman (2014) find using 1970–2010 national balance sheets that in the eight most developed countries the wealth to income ratio first decreased from a very high level of about 600-700% in the eighteenth and nineteenth centuries to around 200–300% in 1970 and since then the ratio has been rising to around 400–600% in 2010. They identify lower population and productivity growth, as well as a recovery in asset prices, as reasons for the recent surge of wealth to income ratios.

Roine and Waldenström show for Sweden (Roine and Waldenström, 2009) and for 10 developed countries (Roine and Waldenström, 2015) that wealth inequality has been falling in the 20th century up to 1980 excluding Switzerland and perhaps the United States where inequality has remained fairly unchanged. They observe that the fall of wealth inequality is mainly caused by the decline of the wealth share of the top 1 percentile. Around 1980, a turning point was reached in most of the countries studied, after which the wealth shares of the richest percentile rose again. The authors attribute this development of wealth inequality to among other factors to globalisation, changes in capital taxation and co-dynamics between income and wealth. In the more recent past, they identify the development of real estate and stock prices as important factors responsible for the fluctuations in wealth shares, as real estate is owned by broader segments of the population, while stocks are quite concentrated in the hands of the wealthiest.

For the last 100 years in the United States Saez and Zucman (2014) estimate wealth through capitalisation of income recorded for individual taxpayers and observe that the development of wealth inequality resembles a U-shape with high levels of inequality in the beginning of the 20^{th} century followed by a decline until the 1980s and with a clear increase since then. The authors assess that the wealth inequality increase in the 30 years between from 1980s to 2010s is primarily caused by the upsurge in the 0.1% share of the wealthiest which rose from 7% in the 1970s to 22% in 2012 and is predominately attributed to the growth of income inequality in combination with a widening of saving rate differentials between different segments of the wealth distribution. Additionally, they evaluate that the dissimilar composition of wealth portfolios for wealthy and the rest of the population can have effects on the wealth inequality estimates, especially in the short term.

A critical assessment of these results comes from Kopczuk (2015) who investigates the development of wealth in the United States using various methods including surveys, the estate multiplier technique and the capitalization method. Until the 1980s Kopczuk (2015) arrives to similar results – that wealth inequality was decreasing – using the different methodologies. From the 1980s on the estimates from the capitalization method diverge from the other two approaches as it indicates a steep increase of the wealth share of the top 1% and 0.1% while the results using the estate multiplier and survey data indicate at most a modest increase, if at all a significant one. He therefore cannot draw an unambiguous conclusion about the wealth development in the last three decades and recommends caution in the use of the capitalization method.

Fagereng et al. (2016) share the criticism of the application of the capitalization method and demonstrate, by comparing Norwegian wealth data obtained from wealth tax with generated wealth data derived using the capitalization method, that the heterogeneity of asset returns and their sensitivity to the level of wealth causes the capitalization method to overestimate the wealth of the top of the distribution.

In a study covering a more recent period between 2000 and 2014 and including 215 countries Davies et al. (2017) find that wealth inequality has been falling everywhere except China and India from the beginning of the millennium until the global financial crisis in 2007 and has been rising since then everywhere except North America. Davies et al. (2017) do not investigate the causes of inequality but supply a substantial dataset covering the vast majority of countries in the world and an estimated 95% of global wealth. They asses that 79% of global wealth inequality is inequality between not within countries. Furthermore, within inequality varies significantly among countries with for instance the USA, China, or Russia where the wealthiest 10% possess more than 70% of total wealth while the top decile in Slovakia, Belgium, or Japan own less than 50%.

The aim of the studies presented above is generally to obtain a clearer picture of the development of the wealth distribution over a greater period from the past till today rather than analysing the underlying causes of the development. Until recently, a pure empirical approach for investigating the causes of wealth inequality was not possible due to a lack of data, which is why theoretical models - usually quantitative models with a general-equilibrium and heterogeneous agents - are the most common approach.

Typically, these are incomplete market models in which households are identical at the beginning, i.e. exposed to the same stochastic income and skill process. Through differences between households in the desire to save and the realisation of income and skill shocks, the households are heterogeneous in the end. The model inputs are usually matched with real world data and the model calibrated so that the resulting wealth distribution equals the one observed in the real world (Cagetti and Nardi, 2008).

Following Cagetti and Nardi (2008) the models can be roughly divided into three groups: models with infinitely lived agents, life-cycle or overlapping generation models and models with a combination of the two preceding. The key idea of all the model specifications is that households want to optimize their consumption over time and therefore save to insure themselves against possible shocks. Life-cycle models add to the motivation for households to save during their working life in order to finance consumption in their retirement. These basic models can be extended to incorporate more motives for saving, like for example a bequest motive, or additional shocks to better match the real-world wealth inequality data.

In a metastudy from 2017 comparing various life-cycle model specifications Nardi and Fella (2017) notice that entrepreneurship, medical expenses and transfers of bequests and human capital from parents to children are the main drivers of wealth inequality. They remark however, that the effects of these variables were investigated usually in isolation so that the informative value could be improved by a model incorporating them jointly.

Scholz and Seshadri (2007) investigate how children contribute to wealth inequality in a life-cycle model and show that young households with a low income are credit constraint if they have children and therefore lack an incentive to save, which leads to higher wealth inequality.

Utilizing a life cycle model, Huggett et al. (2011) examine the importance of initial conditions, defined as at the age of 23, and assert that initial conditions are a more important determinant of lifetime income, wealth and utility than shocks during working life. From the initial conditions evaluated, they show that human capital has a greater impact than learning ability or wealth.

Blandin and Peterman (2019) emphasise the importance of the assumption about human capital accumulation in a life cycle model, whether it is Learning-By-Doing, Learning-Or-Doing or exogenous accumulation. Depending on which of the three learning processes is assumed, different levels of capital taxes are optimal, and this influences the resulting wealth distribution.

In an overlapping generation model Nardi and Yang (2016) identify parental background as a crucial determinant of wealth inequality through the transfer of human and physical capital. Since the latter is only relevant for the wealthiest, they demonstrate that wealth inequality decreases with an increasing inheritance tax rate.

Through the use of alternative approaches in three separate studies - an overlapping generations model (Benhabib et al., 2011), a standard model with infinitely lived agents (Benhabib et al., 2015), in models of earnings inequality and precautionary savings (Benhabib et al., 2017) - Behabib et al. identify that the decisive factor for wealth inequality is capital income risk.

Combining life-cycle and dynastic features Kaymak and Poschke (2016) accentuate that rising income inequality and reforms of the tax and transfer policies in the United States explain a major share of the increase in wealth inequality in the US in the last decades. Another aspect which should be taken into consideration are demographic effects.

Flavin and Yamashita (2002) highlight the importance of housing as an asset and its special features as ownership of real estate differs from other assets, such as bonds, because the own use of the real estate can determine the demand, which does not necessarily have to correspond to the level in a portfolio optimization. According to the data, young homeowners tend to have high credit burdens, which decrease over the lifetime. This affects the portfolio decisions of households over their lifetime, which, with a large enough generation cohort, such as the baby boomers, can have an impact on asset prices and thus on wealth distribution estimates.

Criticism of a key property of life-cycle models comes from Hendricks (2007) who observes that the assumed link between lifetime earnings and retirement wealth in the model is closer than in reality. This causes too large wealth differentials between high- and low-income households and too small variances among households with akin lifetime earnings compared to the real-world data. According to Hendricks (2007) variations in desired consumption growth rate, dissimilarities in discount factors and rates of return can improve the model mimic reality.

Another comprehensive critique of life-cycle models comes from Bernheim et al. (2001). They test key assumptions of the models based on changes in consumption behaviour as people enter retirement and find that variations in pure time preference rates, survival probabilities, income uncertainty, preferences for precautionary savings and inheritance motives cannot explain the observed pattern in the data. Based on the dynamics of the underlying data, Bernheim et al. (2001) conclude that households do not save in a rational and farsighted manner as assumed in life-cycle models, but rather in an estimative and near-sighted way.

Piketty's Capital in the twenty-first century is without doubt the literature concerning wealth inequality which received the most public attention in the last decade (Piketty and Goldhammer, 2014). Piketty's core idea is that wealth inequality is caused by a rate of return on capital r which exceeds the economic growth rate g: r > g. Hence people who own substantial shares of capital will receive larger (capital) income gains than people solely relying on labour income, which grows in average with the overall economic growth rate g. As a consequence, capital owners will be able to accumulate even more capital which in turn will lead to yet greater returns. Therefore, in Piketty's theory r > g results in a self-reinforcing spiral of capital concentration. Hereby, inheritances play a key role in the concentration of wealth through the passing on of large fortunes, and Piketty predicts that the importance of inheritance will dramatically increase in the future.

Piketty underlies his theory with data from France, the United States, and the United Kingdom and presents that wealth inequality has been falling during the beginning of the 20^{th} century and rising since the 1980s. He accounts the initial reduction of inequality to the destruction of capital through the two world wars, the period of high inflation between the wars as well as a combination of high top tax rates and quick economic growth in the post second world war period which caused r < g. The recent increase in wealth inequality is caused by asset price recovery and by a slowing pace of economic growth, due to a slowdown in population growth and technological advancement. Piketty argues that the period of wealth equalization has depicted an anomaly and that in capitalist societies inequality will constantly increase and lead back to the very high levels of historical inequality seen before the 20th century.

While Piketty's contribution to the data availability on wealth is widely acknowledged, his theory has been criticized by many. Blume and Durlauf (2015) present in a detailed critique numerous shortcomings of Piketty's work. Among their criticisms is that in Piketty's definition of capital, human capital and public capital, such as pensions or health care, are not included. In addition, Piketty does not differentiate between productive and non-productive capital nor does he account for depreciation, which significantly affects the development of capital levels, and he offers no theory for saving behaviour, but assumes simply that all capital income is saved while all labour income is consumed. Due to these reasons and other points of critique Blume and Durlauf (2015) reject Piketty's theory for wealth inequality.

Jones (2015) expresses milder criticism, admonishing above all that Piketty simplifies many complexities, such as the reciprocal relationship between the growth rate and the capital gains rate or the assumption of over time constant saving and depreciation rates.

Mankiw (2015) argues that from the capital return rate (r) consumption, the allocation of wealth inheritance between several heirs, inheritance taxes and capital income taxes must be deducted before it can be compared to the economic growth rate (g). Piketty's self-reinforcing wealth concentration spiral would only materialize if the capital return rate (r) minus all the aforementioned factors would still be greater than the economic growth rate (g). Mankiw computes that for the US this is nowhere near the truth and unlikely to occur in the foreseeable future. Furthermore, Mankiw (2015) disputes Piketty’s call for capital taxes by showing that in a neoclassical growth model capital taxes would lessen capital accumulation, labour productivity and wages. Instead he advocates progressive consumption taxes as a more effective method to reduce inequalities.

Inheritance is an important element for wealth concentration in life-cycle models and in Piketty's theory. In order to examine patterns in wealth transfers in the form of inheritances or gifts, Wolff and Gittleman (2014) use survey data from the US between 1989 and 2007. They generally consider wealth transfers to be relevant because, on average over the period studied, around 20% of households recorded a wealth transfer at a given point in time, corresponding to approximately a fourth of their net worth. However, during this period the share of households receiving a wealth transfer has fallen by 2.5 percentage points and a statistically significant change in the size of the transfers is not observed. Therefore, they reject the hypothesis of a "boom" in inheritance and do not expect it to happen in the foreseeable future.

The authors also show that the probability and the absolute value of wealth transfers rise with education, age, income and wealth levels. However, inheritances and gifts increase the wealth of poorer households by a higher percentage than that of the rich, which effectively raises the poor's share of wealth more than that of the wealthy, thereby reducing wealth inequality. Thus, Wolff and Gittleman (2014) find evidence in their study that contradicts the usual assumptions that the importance of wealth transfers has grown lately and that they reinforce wealth inequality.

In addition to life-cycle models and Piketty's theory, there are various approaches to analyse and understand wealth inequality. Bach et al. (2015) explore whether the rich systematically realize higher returns on invested capital than the rest of the population, which if true would widen the gap between rich and poor in the long run. On the basis of Swedish data, they show that the returns on the wealthy's financial assets are indeed around 4% p.a. higher than for the median household. However, this spread is attributable to a greater willingness to take risks by the wealthy rather than to potential superior investment capabilities.

Correspondingly, Fagereng et al. (2016) observe higher returns for wealthier households in Norway but do not investigate the underlying reasons for the disparities.

Using a combination of English administrative and survey data, Bozio et al. (2017) examines whether the rich save more. His result suggests that the higher the lifetime earnings, the greater the share of private savings (measured in wealth to lifetime earnings ratio). When total wealth (including state pension) is used instead of private wealth this linear relationship changes, due to the redistributive nature of the British pension system, and for total wealth the households in the middle of the earnings distribution have the lowest wealth ratio while the 20% of the richest still build up the most wealth. Additionally, the author deduces that education and numerical abilities (being able to calculate compound interest rates) are positively correlated to greater wealth while they find no significant effects of health, expectations of survival or children.

Mares et al. (2018) are the first, to my knowledge, to examine factors of wealth inequality at the cross-country level. Their sample covers 73 countries and they employ Bayesian averaging and data from the newly available Credit Suisse Global Wealth Reports for their wealth inequality variable (Gini coefficient of wealth). Of the 37 different economic, financial, political and social determinants studied, only seven are significant and explain about half of the variation in wealth inequality.

Their results suggest that globalization and war increase inequality while income redistribution and education reduce it. The other three significant variables are all relate to finance and the authors emphasize the important and complex influence of finance on wealth inequality. This is reflected in the fact that they find that large financial markets increase wealth inequality, while better access to finance and more efficient financial intermediaries reduce it. In addition, the authors examine the overall effect that financial development has on wealth inequality by using averages of their financial development measures and find that overall financial development reduces wealth inequality. They therefore recognise that different aspects of finance can have varying implications for wealth inequality, but overall financial development appears to decrease inequality.

In contrast to wealth inequality, income inequality is substantially better understood as the data availability is considerably greater. The majority of the academic literature suggests that income and wealth inequality are strongly, albeit not perfectly, correlated. Hence, a brief analysis of income inequality literature is advantageous for understanding wealth inequality.

According to the findings of Alvaredo et al. (2013), income inequality in the USA rose sharply between 1976 and 2011. During this period, the gross income share of the top earners increased from 9% to 20%. Similar trends can also be observed in other developed countries, even if they are not quite as strong. The authors attribute this development to four reasons: the reduction of top tax rates, a labour market where salary is negotiated and where the reduction of top tax rates may result in managers diverting energy to increase their pay at the cost of business growth and employment, the comeback of capital income and the link between capital income and labour income, where wealthy families yield access to well-paid jobs or even force family members to accept them and thereby become top earners.

Using a panel data approach and including 138 countries between 1960 and 2008, Jauch and Watzka (2016) find that financial development, proxied by the ratio of private credit to GDP, enhances income inequality. In their study, financial development is by far the most decisive variable for explaining differences in income inequality.

Similarly, Bittencourt et al. (2019) conclude on the basis of a panel data study for the US states between 1976 and 2011 that a higher level of financial development leads to increased income inequality. They measure financial development using the ratio of per capita stock market wealth to per capita income.

In a comprehensive study for the IMF, Dabla-Norris et al. (2015) determine that financial depth, technological progress, globalisation as well as less regulated labour markets increase income inequality. On the other hand, improved access to education and healthcare and effective social policies help to reduce the earnings disparities. In their broad sample of nearly 100 countries from 1980 to 2012, the authors highlight the fact that the impact and strength of the drivers of inequality differ between developed and less developed countries. For instance, financial depth has a strong impact on both gross and net income inequality in middle- and low-income countries, while the effect on gross income inequality is markedly lower in developed countries and even negative for net inequality. Dabla-Norris et al. also evaluate the consequences of income inequality and reveal that with a greater share of income to the top 20% income earners economic growth decreases, while it increases when the share of income of the bottom 20% rises.

In a model with infinitely lived agents Pástor and Veronesi (2016) assume that entrepreneurship generates income inequality, and note that differences in ability, participation in the stock market and investment risk are contributory factors. They also find that with the tax rate of a redistributive tax, income inequality initially rise as well as the level of stock prices, but decline eventually with the tax rate.

To summarize my literature review I find that, in a number of developed countries for which such far-reaching data is available, wealth inequality was observed to have decreased since World War II until the 1980s and since then it has been rising. For a substantially larger sample of countries, the overall trend in the new millennium appears to have been a reduction in wealth inequality until 2007 and an increase since then.

A variety of factors have been identified as influencing the level of wealth inequality. A major factor is saving, as a central component of many models, but also in its real-world approximation as financial development. In addition to saving, the most frequently cited factors are income inequality, disparities between human and physical capital, and redistributive policies, which can take various forms, such as higher top and inheritance tax rates or public welfare systems. The role of physical capital is multifaceted, since not only variations in the size of wealth ownership, but also in the composition of physical capital for different segments of the wealth distribution, as well as in the returns realised from it, are considered to be determining factors. Further factors, although not as frequently mentioned, are entrepreneurship, globalisation, war, demographic trends, technological progress and others.

My literature review provides insights into both the development and drivers of wealth inequality. For my empirical analysis, I utilise independent variables that are consistent with the findings of the predominantly theoretical literature on wealth inequality presented here.

3. Data

The quantity and quality of data on wealth is still low compared to other economic variables like income and GDP. The reasons are the difficult observability of wealth, as flow variables like GDP are easier quantifiable than stock variables, and the question of measurement.

Various methodologies have been utilized to estimate the levels and distribution of wealth each with its own advantages and shortcomings. The most popular methods are to use wealth tax data (if available), data from surveys, the estate multiplier method, which employs inheritance tax return data to calculate wealth for the wealthiest part of the population, the capitalization method, for which capital income tax data is used to deduce the value of the underlying assets and finally so-called “rich-lists” like the probably most famous one issued by Forbes. These diverse approaches can either be used individually or in combination to estimate wealth.

Besides the choice of methodologies differences in wealth estimates exist for the unit of the wealth holder, what to include as wealth and how to correctly evaluate the wealth. Thus, whether wealth is calculated per individuum, per adult person, per household or per family yields potentially different wealth estimates as well as if for instance public pensions, consumer durables, human capital and indirect wealth ownership (i.e. through foundations or trusts) are included or not. Whether to evaluate wealth at market prices or balance sheet prices, before taxes or after taxes and whether using gross worth or net worth (gross worth minus liabilities) alters the results as well.

Discussing the advantages and disadvantages of each the different approaches would be a work on its own and I therefore will not pursue this and instead will focus in the forthcoming on the critical assessment of the methodology that underlies the wealth data I use.

In my thesis I utilise wealth data from the Credit Suisse Global Wealth Report databooks (Shorrocks et al., 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018)[5]. The databooks, which are published each year since 2010, contain data covering 172 countries in the period between 2000 and 2019 with annual observations[6].

The wealth data provided comprises different series of figures on wealth, ranging from absolute variables such as the average wealth, financial and non-financial wealth, and debt per capita to relative variables such as the Gini coefficient of the wealth distribution or the share of wealth owned by certain segments of the population. Since I examine inequality in my study, I will focus on relative variables, especially on the Gini coefficient of wealth. The Credit Suisse Global Wealth Report databooks data is compiled according to the methodology described by Davies et al. (2017) which I discuss below.

Davies et al. (2017) apply a definition of wealth as net worth, i.e. the market value of financial assets plus non-financial assets minus liabilities. Thereby they include private pension wealth but exclude public pensions and they measure wealth on the unit of an individual adult (i.e. 20 years and older).

For 33 countries, including all major high-income countries and the developing countries with the largest populations - China, India and Indonesia - direct data on the distribution of wealth is available to Davies et al. (2017). The direct data is based mostly on household sample surveys, but also on wealth tax records or register data containing both wealth tax records and additional information.

They refine the survey data as the wealthiest segment of the population tends to be under-represented in surveys and the wealthiest individuals are often omitted altogether. The refinement consists of using the number of billionaires reported by Forbes magazine for the respective countries each year to match a Pareto distribution to the upper end of the wealth distribution and thereafter replacing the top wealth values with these new estimated values. Then, Davies et al. (2017) re-scale the resulting revised sample for each country to correspond to the mean wealth value. Moreover, they harmonise the data of the countries via the ECB's

[5] The databooks contain detailed data series on wealth and comprise the data collection of Credit Suisse's annually published Global Wealth Reports (Credit Suisse Research Institute, 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018). In my case, I was so fortunate to receive the wealth data personally from Prof. Shorrocks, one of the authors of the Global Wealth Reports.
[6] The first Global Wealth Report in 2010 covered the years from 2000 to 2010, and the subsequent ones the respective following year.

Household Finance and Consumption Survey (HFCS) and the Luxembourg Wealth Survey (LWS) for the countries covered by them.

Davies et al. (2017) find that based on the data of those countries with available direct data a relationship between the wealth distribution and the income distribution can be assumed. They exploit this relationship to estimate, under consideration of the respective geographical region of each country, the wealth distribution for 135 other countries which have data on income distribution but not on wealth distribution.

Countries for which no income inequality data is available are assigned a wealth distribution pattern corresponding to the average (weighted by the size of the adult population) of their respective region and income class. They justify this rather rough estimation by stating that they choose to do this rather than omit countries altogether, which would implicitly assume that their pattern of wealth distribution corresponds to the world average. They argue that verification tests indicated that excluding the roughly estimated countries from the global sample has little impact on the conclusions.

The methodology employed by Davies et al. (2017) has implications for my later results. In particular, three aspects should be mentioned: only for a few countries direct data on the distribution of wealth is available while for the most countries it has been estimated; the key variable for this estimation is income inequality; and public pensions are not included in the wealth definition.

The first point is relevant because I am using more countries than those for which direct wealth data is reported. In those cases where data has been calculated, the probability of endogeneity due to measurement errors is particularly high. I address this possible problem in chapter 6.4.2 ‘Further robustness checks’ by performing a robustness check by conducting a regression on a sample consisting only of countries with direct data. Generally, Davies et al. (2017) appear to have exercised due diligence in the compilation of their data, which is reflected in the various adjustments of their data, hence I can rely on the data to a reasonable extent.

Secondly, since Davies et al. (2017) use income inequality as their core variable to estimate missing wealth inequality data and as income inequality is among the variables I employ, a significant relationship between wealth and income inequality can be expected in my results. However, this should not be a major problem for my overall regression results as income inequality represents only one of many variables I use, so it should not affect the significance of the other variables.

Finally, when interpreting my results, it is important to consider that public pensions are not included in the definition of wealth, so higher inequality in a country with a considerable public pension is less severe in its implications for people than in a country without a significant public pension.

4. Lasso

4.1. Country and year selection

Through Lasso, I am pursuing rather a data mining than theoretical approach to identifying the determinants of wealth inequality. My selection of variables, years and countries is always subject to a trade-off between the incorporation of as much relevant data as possible and at the same time achieving a high degree of balance in my data.

Of the 172 countries and the period between 2000 and 2019 covered by the wealth data I only consider the years between 2000 and 2018 and 129 countries in my baseline regression. I disregard the year 2019 because most of my independent variables only provide data up to 2018 and sometimes even only up to 2017.

Furthermore, I exclude 39 countries which have an (adult) population smaller than one million at the beginning of my sample period in 2000[7], since, by the very nature of my methodology, each state is attributed the same weight in the regression, regardless of its size. The economic systems of small states are often highly dependent on a few factors, such as tourism or their role as tax havens, and are typically not comparable with those of larger countries. Consequently, the inclusion of small countries could bias my results in a way which would make them less meaningful for the majority of people, who do not live in small countries.

In addition, I exclude four more countries - Afghanistan, Eritrea, Turkmenistan, Taiwan - from my baseline regression due to insufficient data. For each of these four nations, data is missing completely for at least 35 of 87 independent variables which I employ for lasso and additionally for some of the remaining variables their data is incomplete. The countries remaining in my sample have a maximum of 18 out of 87 variables with completely missing data.

Following these reductions, it leaves me with 129 countries and 19 annual observations between 2000 and 2018. Moreover, the remaining data set still covers a considerable geographical range as it includes 29 African countries, 21 countries from Asia and Pacific, 36

[7] Size of population based on the data from the Global Wealth Report databooks (Shorrocks et al., 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018).

European, 24 from Middle East and Central Asia and 19 from the Western Hemisphere and is equally diverse with regard to income groups, as it comprises 31 advanced markets, 58 emerging markets and 40 low-income and developing countries[8]. In a robustness check in chapter 6.4.2 'Further robustness checks' I consider all countries, including those I have omitted here.

Please consult Table A1 in the appendix for details on the individual countries and their respective region and income group classification.

4.2. Variable selection

The literature mentions a wide variety of factors that are considered to influence wealth inequality. The most frequently identified drivers include saving and associated differences in portfolio composition, education, income inequality, demographic trends, redistribution, globalisation, and entrepreneurship.

However, since so far only the study of Mares et al. (2018) is comparable to my research, in that it empirically examines wealth inequality at the country level, no benchmark has yet been established as to which variables should be used. As a result, I apply lasso (least absolute shrinkage and selection operator) for fixed effects to select the likely most decisive variables from a multiplicity of potentially meaningful ones.

I consider a total of 87 independent variables, which can be roughly divided into eight categories: financial development, globalisation, education, entrepreneurship, taxes/redistribution/income inequality, political institutions, war or coups and general economic or other control variables. Within each category, my variables represent various aspects of the category or reflect different measurements of the same aspect. More details about the independent variables I include in my lasso estimation can be found in Table A2 in the appendix.

Because of this large number of variables, all factors mentioned in the literature, for which data is available, are included, among them the exact seven variables that Mares et al. (2018) found in their work as the main drivers of wealth inequality. Of the 87 variables, I utilise 82 variables that satisfy the criterion of being at least 75% balanced for my baseline sample, i.e. having data for at least 75% of my observations (over countries and years).

---

[8] Both the classification of countries by region and by income group are taken from those of the International Monetary Fund (International Monetary Fund 2020a).

The excluded five variables are all measures of performance of different features of the stock market and are less than 60% balanced as they are usually only available for higher developed countries. They nevertheless may yield interesting information and I assess their significance in a later robustness check in chapter 6.4.2 'Further robustness checks', in which I only consider developed countries.

Out of the 82 independent variables employed for lasso, 37 are very well balanced (i.e. 95% balanced or more), 31 variables are well balanced (i.e. 85% balanced but less than 95%) and 14 variables are fairly balanced (i.e. between 75% and 85% balanced). The overall sample is well balanced with 93.5%.

4.3. Methodology

Being the second, after Mares et al. (2018), to empirically study wealth inequality using country-level data, no certainty exists as to which variables are meaningful determinants. Out of this model selection problem I employ lasso. The least absolute shrinkage and selection operator (lasso) is a data-driven model that selects the most informative variables under the constraint of sparsity. In its basic form, as originally created by Tibshirani (1996), it has the following features:

$$\hat{\beta} = argmin_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \text{ subject to } \sum_{j=1}^{k} \left| \hat{\beta}_j \right| \leq t \tag{1}$$

Thus, lasso is the optimisation of ordinary least squares subject to the constraint that the aggregate size of the coefficients is limited up to t. Here t is a tuning parameter that controls the shrinkage applied to the estimates. If t_o represents the L1 norm of OLS estimators, then in the case of $t \geq t_0$ the shrinkage will have no effect, but when $t \leq t_0$ the shrinkage will force the estimates towards zero, which often leads to some estimates being zero, thus eliminating variables. Belloni et al. (2016) have extended the initial model and developed lasso models for fixed effects panel model applications by enabling cluster-dependent errors.

Although I cannot know the correct model which my panel data will assume, I choose to use lasso for fixed effects rather than for pooled OLS, because my dataset contains very diverse countries, which makes it probable that country-specific unobserved time-invariant effects are present. My assumption of fixed effects is confirmed later by the appropriate model selection tests[9] run on the sample including the variables chosen by lasso.

Belloni et al. (Belloni et al., 2011; Belloni et al., 2012; Belloni et al., 2016) provide theory-based penalty terms t for lasso both on the assumption of homoscedastic and heteroscedastic errors, and since I have no theoretical knowledge about the nature of my errors, I apply both homoscedastic and heteroscedastic errors in separate instances. The necessary code is available in the form of the lassopack for STATA (Ahrens et al., 2020).

To be mentioned is that lasso and my subsequent fixed effects model are different procedures. Lasso estimates the coefficient values while minimising a penalty term based on the sum of the absolute values of the coefficients, while fixed effects tests the null hypothesis of statistical significance of a coefficient given a smaller model as a background assumption. It is possible that lasso selects variables that are then estimated to be insignificant by fixed effects, as well as that lasso does not select variables that are estimated to be significant by fixed effects. In other words, lasso selects influential variables under a parsimony constraint and not all variables that are potentially influential in fixed effects.

Furthermore, it is important to note that lasso, being data driven, can identify the variables that have the greatest statistical influence on the dependent variable under the constraint of the shrinkage, but it cannot determine the causality of the relationship. Therefore, for the pool of variables available for lasso selection I only consider independent variables that should have a causal relationship with the dependent variable based on the theory obtained from my literature research.

4.4. Results

The two different specifications of the penalty term for lasso which I employ, homoscedastic lasso and heteroskedastic (robust) lasso, yield similar results. Homoscedastic lasso selects eight variables and heteroscedastic lasso selects six variables out of the pool of 82 variables. There are five variables that are selected equally by both specifications. Table 1 lists the total of nine unique variables selected by the two lasso specifications, along with details regarding which specification selected them. In the appendix more detailed results for homoscedastic lasso are presented in Table A3 and for heteroskedastic lasso in Table A4.

[9] Breusch and Pagan Lagrangian multiplier test and Hausmann test for fixed or random effects model

Table 1: Lasso variable selection

Variable	Homoscedastic lasso selected	Heteroskedastic lasso selected	Source of the Variable
Exports of goods and services (% of GDP)	Yes	Yes	(The World Bank, 2020c)
Labour force participation rate, total (% of total population ages 15+)	Yes	Yes	(The World Bank, 2020e)
Financial market depth index	Yes	Yes	(International Monetary Fund, 2020b)
Education index of the UN	Yes	Yes	(United Nations, 2020)
Domestic credit to private sector (% of GDP)	Yes		(The World Bank, 2020b)
Human capital index	Yes	Yes	World Penn Table (Feenstra et al., 2015)
Globalisation index economic dimension	Yes		ETH Zurich (Gygli et al., 2019)
Business regulations index	Yes		(Fraser Institute, 2020a)
Control of corruption: estimate		Yes	(The World Bank, 2020a)

Notes: This table lists the nine individual variables selected from either lasso specification, lasso for fixed effects assuming once homoscedastic and once heteroscedastic errors, including from which lasso specification they were selected respectively. In addition, the sources of the variables are provided.

It is reassuring for the validity of the variables selected that the two different specifications selected similar parameters from the large pool of possible variables. Since I have quite an extensive data set, I consider all nine variables in the following, even the four that were not equally selected by both lassos. In this way, I intend to include as much information as possible in my models.

For my baseline regression I choose to include only seven of the nine variables selected by lasso. The reason is that for both of the two variables I omit, there exists a corresponding variable which is included in my baseline regression which represents the same indicator. Specifically, the education index of the UN (United Nations, 2020) and the human capital index of the World Penn Table (Feenstra et al., 2015) each measure education. Exports of goods and services (The World Bank, 2020c) and the economic dimension of the Globalisation index

(Gygli et al., 2019) furthermore are both indicators of globalisation related specifically to the economic dimension.

Their respective similarity is reflected in the correlation between the two variables within each pair of variables: High correlations are present prior to the incorporation of fixed effects, as well as high VIF values following pooled OLS regression and, most importantly, high correlations of the parameters following fixed effects model regression[10]. For reasons of parsimony and reduction of multicollinearity, I thus utilise only one of the variables in each pair.

I choose the education index of the UN for education and exports of goods and services for globalisation. My decision is based on two criteria: lasso selected each of these variables before their respective equivalents, i.e. according to lasso these variables are more meaningful, and the variables chosen are more balanced, i.e. more data is available. In chapter 6.4.2 'Further robustness checks' I utilize the omitted variables in a robustness check, substituting the respective variables from each pair with the corresponding omitted variables.

In chapter 5.2, after a descriptive analysis of my dependent variable, I discuss the properties of the seven independent variables I include in my baseline regression in more detail.

5. Descriptive Analysis

Prior to proceeding with my baseline model estimation, the fixed effects model, I analyse both my dependent variable, the Gini coefficient of wealth, and my independent variables.

5.1. Gini coefficient of wealth

The Gini coefficient of wealth is a statistical indicator of wealth inequality in the population of a country. The coefficient measures the distribution of wealth between members of a population and can assume values between 0 and 1 or, as here, between 0% and 100%. Hereby, a Gini value of zero denotes a perfectly equal distribution of wealth within a population and a coefficient of 100 stands for perfect inequality, i.e. a situation in which one person of a population owns all the wealth, while all the others possess nothing. In general, the higher the value of the Gini coefficient, the greater the wealth inequality. Thus, the Gini coefficient is a relative measurement that represents the distribution of wealth rather than the absolute level of wealth. Since the computation of the Gini coefficient is unaffected by how populous or wealthy

[10] The respective values are: for the educational pair: pre fixed effects: 0.95, VIF: for both variables above 10, post fixed effects: -0.83; for the globalisation pair: pre fixed effects: 0.63, VIF: globalisation economic dimension more than 5 and exports of goods and services: 1.8, post fixed effects: -0.45

the single countries are, the wealth Gini makes wealth inequality comparable between countries that have highly different characteristics.

The wealth inequality data is fairly new and therefore relatively unknown. For this reason, I perform a rather detailed basic descriptive analysis of my data prior to my regressions.

As shown in Table 2, the wealth inequality value averaged over years and countries is 72.89 and has a relatively small standard deviation of 6.96. The low variance is reflected in the absence of many extreme values: only three countries recorded a Gini value higher than 90 in any given year - Russia 2002 to 2003, the Netherlands 2015 to 2018 and Lebanon 2007 - and only one country had a Gini lower than 55 - Slovakia from 2000 to 2018.

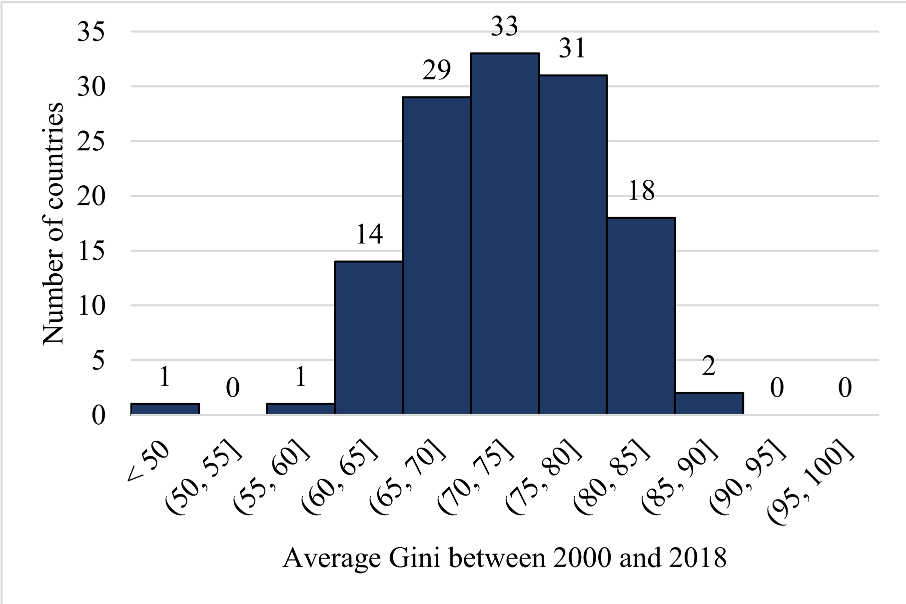
Table 2: Gini of wealth descriptive values

	Wealth Gini
Mean	72.89
Min	46.71
Median	72.53
Max	91.56
Standard deviation	6.96
Between std	6.63
Within std	2.19
Degree of balance	100%

Notes: std as in Between std and Within std stands for standard deviation

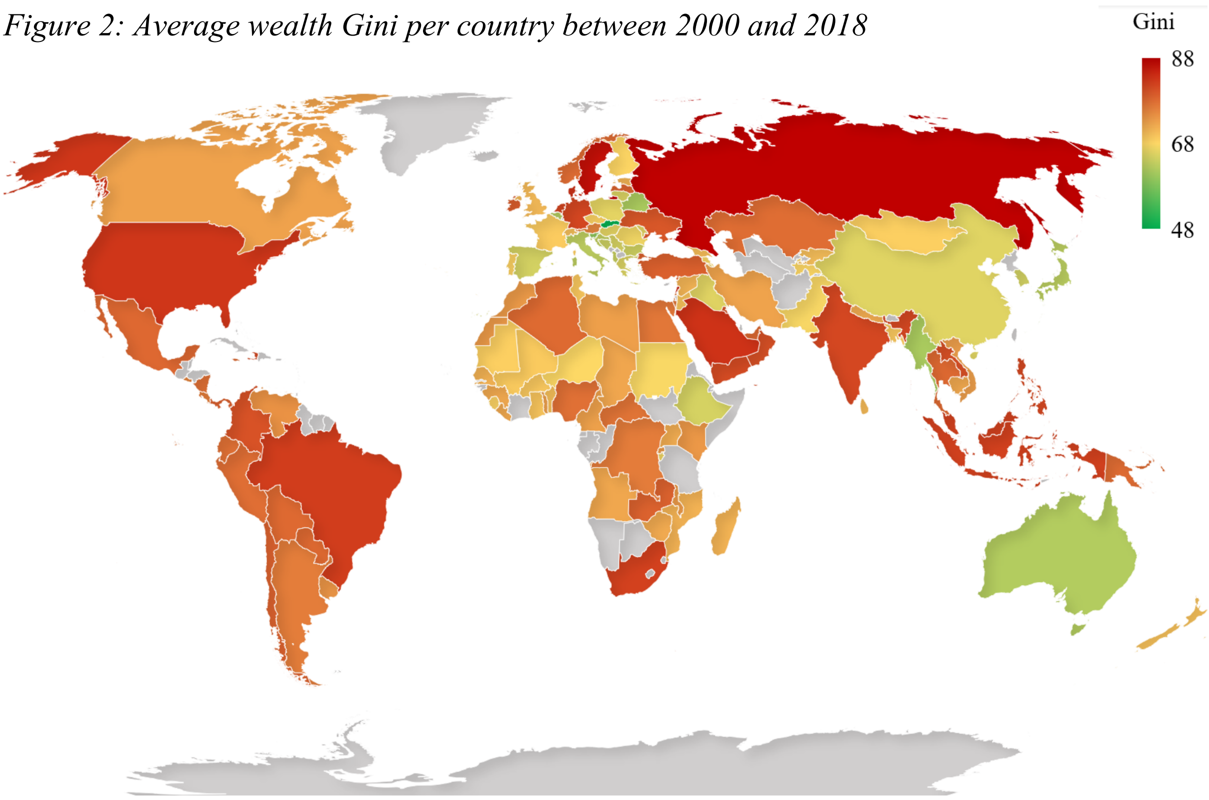
The between variance, i.e. the variance capturing the differences between countries, is substantially larger than the within variance, i.e. the variances capturing changes in given countries over time. Figure 1 illustrates the between variance, i.e. the distribution of the average Gini values of the countries averaged over the years. It is apparent that the vast majority of countries have Gini values between 65 and 85. Figure 2 provides a graphical presentation of the average Gini coefficient between 2000 and 2018 for the 129 countries included in my baseline sample.

Figure 1: Distribution of countries by average wealth Gini between 2000 and 2018



Notes: The figure shows a histogram of the distribution of countries according to the value of their average Gini coefficient of wealth between 2000 and 2018.

Figure 2: Average wealth Gini per country between 2000 and 2018



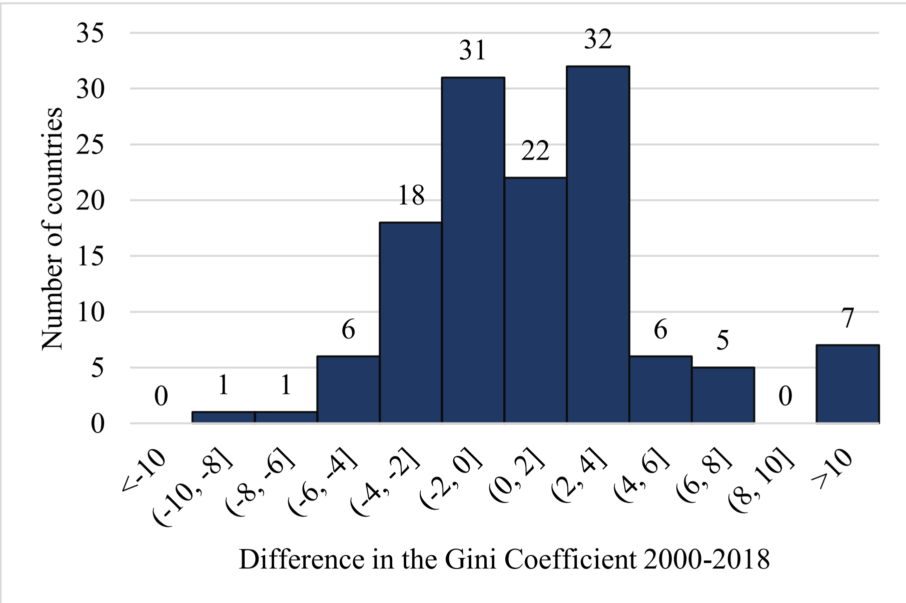
Notes: The figure shows a world map where countries are coloured according to their average Gini coefficient of wealth between 2000 and 2018. The colour scale corresponds to those of a traffic light: low wealth inequality is indicated by green, medium wealth inequality by yellow and high wealth inequality by red, and there are many nuances in between. Countries coloured grey are not included in my baseline regression sample of 129 countries.

The development over time of the mean Gini value of all countries exhibits a slight upward time trend, with the average Gini in 2000 being 72.47 and increasing by 1.08 points reaching 73.55 in 2018. This gradual increase is not strictly continuous, as both increases and decreases of the mean Gini occurred over the observation period. The average Gini reached its minimum value in 2002 with 71.67 and its maximum in 2013 with 74.26, after which it fell again and rose again slightly by 2018.

A closer inspection of the developments between 2000 and 2018 in the individual countries reveals a diverse picture. As described above, the average Gini coefficient for wealth inequality increases by 1.08 points during this period. Considering that the standard deviation is 4.08 points indicates that the significance of the trend is weak. In fact, the median growth only amounts to 0.66 points. In most countries, the change between 2000 and 2018 is marginal: in 103 nations (80%) the Gini coefficient has risen or fallen by only four points and in 29 countries (22.5%) within one point, meaning inequality there has remained virtually unchanged.

Figure 3 illustrates this fact, clearly showing that most of the changes are close to zero, with only a few exceptions showing a change of more than four points.

Figure 3: Distribution of countries by differences in the wealth Gini between 2000 and 2018



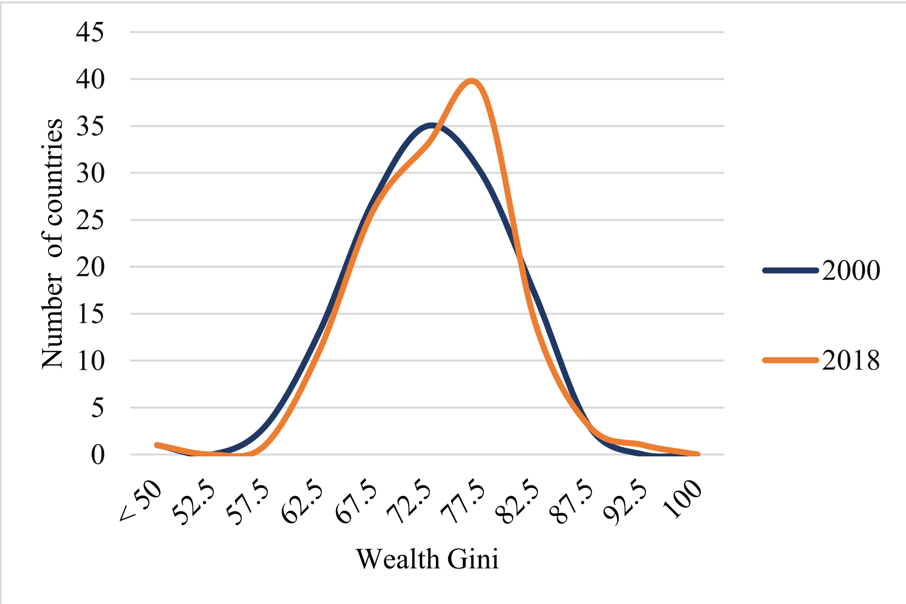
Notes: The figure shows a histogram of the distribution of countries according to the value of the difference of their Gini coefficient of wealth between 2000 and 2018.

In 72 countries (55.8%) inequality increased between 2000 and 2018, while in 57 countries (44.2%) inequality decreased. However, more countries recorded a steep rise in inequality than countries saw a sharp reduction. For instance, in seven countries in my observation period inequality increased by more than 10 points, while there is no country where it fell by 10 points.

The Ukraine saw the strongest increase with 19.5 points, while Libya achieved the largest reduction with 8.5 points.

The development is further visualised by Figure 4 on which two density plots based on the Gini coefficient from 2000 and 2018 are placed on top of each other. Fundamentally, they are similar but with a slight shortening of the left tail and a shift of the maxima from around 72.5 to around 80 points between 2000 and 2018. It is worth noting that the two most populous countries - China and India - are both in the top 10 countries with the largest increases of 11.6 and 7.2 points respectively. This implies that although no strong trend can be identified at the country level, inequality has nevertheless increased for a substantial share of people worldwide within their countries.

Figure 4: Distribution of countries by wealth Gini 2000 vs 2018



Notes: *The figure shows two density plots visualising the distribution of countries according to the value of their Gini coefficient of wealth in 2000 (dark coloured line) in comparison to 2018 (light coloured line).*

When examining the level and development of wealth inequality by (IMF) region, it can be found that Europe has the lowest inequality with an average of 70.28 points and the Western Hemisphere the highest with a Gini coefficient of 77.95. The Middle East and Central Asia, Africa and Asia and Pacific are between these extremes with around 73 points each. Europe and Africa experience the largest increases in wealth inequality between 2000 and 2018, each exceeding 2 points, while Asia and the Pacific show a minor rise of less than one point and the Middle East and Central Asia and the Western Hemisphere both record reductions of less than one point.

A breakdown into (IMF) income groups shows that advanced markets have the lowest average inequality with a Gini coefficient of 71.1 points, followed by low-income developing countries with 72 points and then emerging markets with 74.5 points. In all three income categories, inequality rises over the observation period with the smallest increase of 0.8 points in emerging markets and a plus of 1.3 points in both advanced markets and low-income countries.

This preliminary analysis of my data shows that a large part of the overall variance in wealth inequality is between variance of which only a minor share or nothing at all can be captured by my fixed effects models. However, I find that a lot of interesting information is contained in the within variation, since the development of wealth inequality in different countries over time is characterised by a high degree of diversity.

Based on this data, I cannot identify a clear trend of an increase in wealth inequality, at least at the country level.

5.2. Independent variables

Subsequently I will briefly describe the seven independent variables which I will employ in my baseline regression. Table 3 provides an overview of the properties of my independent variables. The values are computed across all countries and years, i.e. without country-specific demeaning as in fixed effects. An exception to this is the between standard deviation, which represents the variance between the countries, i.e. with the years dimension reduced, and the within standard deviation which represents the variance within the countries over the years, i.e. with the countries dimension reduced.

Of the seven variables, four are indices: depth of the financial market which is a subindex of the Financial Development Index of the IMF (International Monetary Fund, 2020b), the education index from the Human Development Reports of the UN (United Nations, 2020), the business regulations index which is a subindex of the Economic Freedom Index published by the Fraser Institute (Fraser Institute, 2020a) and the control of corruption estimate from the Worldwide Governance Indicators Database of the World Bank (The World Bank, 2020b). Since these indices partly have considerably varying scales, I conducted a linear transformation in order to standardize the indices on a scale from 0 to 100, thus improving the comprehensibility of the regression results and facilitating the comparison of the impact of the different indices. This linear transformation leaves the functional relationship between my transformed variables and the dependent variable intact, and only affects the size of the estimated coefficients and the standard deviation.

Of the three non-index variables, exports of goods and services (The World Bank, 2020c) and domestic credit to the private sector (The World Bank, 2020a) are measured as a percentage of GDP and labour force participation (The World Bank, 2020e) as the share of employed persons of the total population aged 15 and above. All three are part of the World Bank's World development indicators database (The World Bank, 2020f).

Table 3: Independent variables descriptive values

	Exports of goods and services	**Labour force participation rate**	**Financial market depth**	**Domestic credit to private sector**	**Education**	**Domestic credit to private sector**	**Business regulations**	**Control of corruption**
Unit	In % of GDP	In % of total population above 15 years old	Index from 0-100	In % of GDP	Index from 0-100	In % of GDP	Index from 0-100	Index from 0-100
Mean	39.57	62.65	25.33	57.46	61.63	57.46	62.69	48.14
Min	0.10	36.83	0.00	0.19	11.60	0.19	19.41	15.56
Median	32.60	62.37	10.59	41.29	64.20	41.29	62.99	42.10
Max	228.99	89.05	100.00	257.18	94.60	257.18	94.00	99.40
Standard deviation	27.89	10.89	29.59	47.58	19.41	47.58	13.62	20.91
Between std	26.65	10.77	29.00	45.41	18.96	45.41	12.39	20.69
Within std	8.26	1.88	6.42	16.11	4.36	16.11	5.65	3.53
Degree of balance	97.18%	100%	98.45%	78.50%	99.67%	78.50%	90.98%	94.74%

Notes: std as in Between std and Within std stands for standard deviation

In the following, I conduct a more in-depth analysis of the variables in order to improve the understanding of their potential influence on the wealth Gini.

The financial market depth index is a sub-index of the International Monetary Fund's index of financial development. The IMF's definition of financial markets thereby includes stock and bond markets, and depth is computed by first normalising and then aggregating five economic variables that describe the size and liquidity of the markets[11] (Svirydzenka, 2016). The standard measure of financial market depth - stock market capitalisation - serves as one of the five variables for the construction of the index. The deeper the financial market in a country, the higher the value of the index. The mean value across all years and countries is 25.33 and the standard deviation 29.59.

The greatest depth of the financial markets in any year is attained by Canada in 2017 with the maximum value of the index of 100. Several countries have the lowest possible value of the index at 0, Central African Republic, Haiti and Malawi over the entire period from 2000-2018, Benin from 2001-2018, Syria from 2009-2018 and Libya from 2014-2018. These countries accordingly exhibit the lowest possible depth of the financial markets in the corresponding years.

The education index is based on the average years of schooling and the expected years of schooling. In this context, the higher the value of the education index, the longer the population enjoys education. The mean value is 61.63 with a standard deviation of 19.41. In any given year and country, Germany scores the highest value with 94.6 in the years 2016 to 2018, i.e. has the longest average education, while Niger has the shortest mean education with the lowest value in the education index of 11.6 in the year 2000.

The business regulations index is composed of several sub-indices[12], which describe the regulatory business environment in a given country. Thereby holds that the higher the value of the business regulations index for a country, the more business-friendly the country. In my sample, the business regulations index reaches its maximum value with a value of 94 for Singapore in 2017 as the most business-friendly country in any year, and its minimum value of 19.41 for Venezuela also in 2017 as the country with the most unfavourable regulations for enterprises. The mean value amounts to 62.69 and the overall standard deviation to 13.62.

The control of corruption index is based on the perception of the degree by which public power is exploited for personal benefit[13]. The higher the index value, the stronger the perceived control of corruption, ergo the lower the felt level of corruption. The mean value of the control

[11] The five variables are: Stock market capitalization to GDP, stocks traded to GDP, international debt securities of government to GDP, total debt securities of financial corporations to GDP, total debt securities of nonfinancial corporations to GDP

[12]Composed of sub-indices: Administrative requirements, regulatory burden, starting a business, impartial public administration, licensing restrictions, tax compliance

[13] This includes the "capture" of the state by elites and private interests, as well as minor and major forms of corruption.

of corruption index equals 48.14 with a standard deviation of 20.91 and a maximum value of 99.4 recorded for Denmark in 2006 which contrasts with the minimum value of 15.56 registered for Haiti in 2003.

Exports of goods and services to GDP as a measure of outward orientation has a mean of 39.57% and a standard deviation of 27.89 percentage points. Myanmar in 2011 has the lowest outward orientation with 0.01% exports as a share of GDP and Singapore in 2008 the highest with 228.99%.

Domestic credit to the private sector as a percentage of GDP is a conventional indicator for assessing the depth of financial institutions. Here the average value is 57.46% with a standard deviation of 47.58 percentage points. Slovenia has the minimum recorded value of only 0.19% domestic credit to GDP in 2004, while Mauritania registered the highest depth of financial institutions in 2009 with 257.18%.

The average labour force participation rate is 62.65% and has a standard deviation of 10.89 percentage points. The highest level of unemployment over my sample was recorded in Yemen in 2014 with a labour force participation rate of only 36.83% and the lowest unemployment was reported in Madagascar in 2010 with a participation rate of 89.05%.

A common feature of all independent variables is that the between variance is considerably larger than the within variance. The ratio of between variance to within variance ranges from slightly more than double in business regulations to almost six times as large in the control of corruption index. Given that my data set covers a wide spectrum of diverse countries but only a relatively short period of time this is not surprising. My main model specification, fixed effects, primarily describes within variance, ergo a lot of information cannot be accounted for in my regressions. However, it is also evident that quite a substantial amount of variance is still present in the within variance.

My baseline regression is quite well balanced at 94.9% across my seven variables. Domestic credit to private sector is by far my least balanced variable with 78.5%, all the other variables are at least 90% balanced. Overall, nine nations each have completely missing data for one variable, which is for six of them domestic credit to private sector. 60 countries have at least one variable where gaps in the data occur. Finally, 66 countries have perfect data availability for all variables, so they are 100% balanced.

6. Baseline regression

6.1. Methodology

In the following I will shortly discuss the main method which I employ in my thesis: the fixed effects model.

My estimated model assumes the following form:

$$Gini\ Wealth_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it} \qquad (2)$$

The dependent variable $Gini\ Wealth_{it}$ represents wealth inequality measured as Gini coefficients of wealth in year t for country i. Furthermore, α_i corresponds to the unobserved time-invariant country-specific effect and ε_{it} to the error term. The vector X'_{it} – with i being the cross-sectional and t the time dimension – contains the seven independent variables which I employ in my baseline regression, specifically exports of goods and services, the financial market depth index, domestic credit to private sector, labour force participation rate, the control of corruption index, the education index and the business regulations index. The β coefficients in a fixed effects model should be interpreted as follows: For a given country, if the independent variable X changes over time by one unit, while all other factors are held equal, then the dependent variable Y increases or decreases by β units.

6.2. Results

Both the null hypothesis of the Breusch and Pagan Lagrangian multiplier test for the selection of pooled OLS versus random effects model and the null hypothesis of the Hausmann test for fixed or random effects model are rejected at the highest level. Thus, fixed effects is the efficient and consistent estimator for my data. Hence, I estimate my baseline regression with fixed effects and employ robust standard errors for reliability against heteroskedasticity. The null hypothesis of the F-statistic in my fixed effects estimation is rejected with a value of 5.32 and a p-value of 0 at the highest significance level, suggesting that the model including independent variables provides additional information over a model with only an intercept.

The results of my baseline regression are presented in Table 4. Six of the seven independent variables are significant with only domestic credit to the private sector being insignificant. Labour force participation rate, education and business regulations are highly significant at the 1% level, while exports of goods and services, the depth of the financial market and control of corruption are significant at the 5% level. A higher labour force participation rate and greater control of corruption exhibit a negative correlation with wealth inequality while a higher exports to GDP ratio, i.e. greater globalisation, deeper financial markets, a higher value in the education index, i.e. more education, and a more business friendly regulatory environment are associated with increased wealth inequality.

Table 4: Regression results baseline regression

	Dependent variable:
	Wealth Gini
Exports/GDP	0.038^{**} (0.017)
Labour force participation	-0.232^{***} (0.057)
Financial market depth	0.037^{**} (0.018)
Domestic credit	0.002 (0.007)
Education	0.068^{***} (0.025)
Business regulations	0.068^{***} (0.021)
Control of corruption	-0.069^{**} (0.031)
Observations	1,698
Within R^2	0.150
F Statistic	5.32^{***} (df = 7; 116)
Note:	$^{*}p<0.1, ^{**}p<0.05, ^{***}p<0.01$

The results are overall consistent with economic intuition and the limited insights already presented in the literature. The beta coefficients in a fixed effects model should be interpreted according to the following concept: for a given country, a variation of the regressor in time by one unit leads ceteris paribus to a change in the wealth Gini of beta size. The subsequent discussion of my results should be understood accordingly.

My results suggest that a one percentage point higher export share of GDP is associated with a 0.038 Gini point increase in wealth inequality, everything else held equal. Simply put, there appears to be a positive correlation between the level of globalisation and wealth inequality.

Similarly, greater depth of financial markets affects inequality, where an increase in depth by one index point is associated ceteris paribus with 0.037 Gini points higher inequality. In the case of this finance variable, more finance implies greater wealth inequality. Equivalent to my results Mares et al. (2018) identified precisely these two variables - exports of goods and services to GDP and depth of financial markets - previously as having an amplifying effect on wealth inequality.

Also linked to higher wealth inequality is the business regulations index, where an increase of one index point leads to a 0.068 Gini points rise in wealth inequality. Since the higher the index value, the more beneficiary the regulatory environment for starting and running a business, a connection can be established between improved opportunities for entrepreneurship and higher wealth inequality.

The education index of the UN, which is based on the average years of education, exhibits a positive correlation with inequality which is slightly surprising. Here an increment of one index point in education is ceteris paribus equated with a 0.068 Gini point increase in inequality. On the contrary, Mares et al. (2018) find an inequality reducing effect of exactly this education indicator in their study. Moreover, it is more in line with economic intuition that higher average education should lead to lower and not higher inequality. Since education is the variable that reflects intuition and past findings the least, the positive relationship could be due to endogeneity, so I will give a special emphasis to it in my robustness analyses.

An increase of the labour force participation rate of one percentage point leads ceteris paribus to a reduction of -0.232 Gini points in wealth inequality. In simple terms, wealth inequality appears to decrease as the share of the population that is in employment rises.

The control of corruption also exhibits a decreasing effect on wealth inequality with -0.069 Gini points ceteris paribus per index point. Ergo, it appears that lower levels of corruption result in lower levels of wealth inequality.

Finally, my second finance variable, domestic credit to the private sector, a measure of the depth of financial institutions, appears not to be a significant factor.

6.3. Potential sources of error

My results could be subject to bias caused by several possible problems. This could render them invalid. Therefore, in the following section, I discuss the most likely potential problems. In addition, I discuss the approaches in how I intend to validate my results and examine how they withstand robust scrutiny.

6.3.1. Multicollinearity

Multicollinearity occurs when two or more of the regressors are highly correlated. In the case of multicollinearity, the coefficients and the statistical interference of the variables concerned are distorted.

In my baseline regression collinearity was partly quite high before the fixed effects transformation according to the correlation matrix of the regressors and high VIF values I obtained after a pooled OLS estimation. However, examining the collinearity between the parameters post fixed effects estimation shows that after the fixed effects transformation no high collinearity is present. Accordingly, multicollinearity should not pose a problem in my estimation.

Nevertheless, as part of my robustness checks, I will assess whether my results are robust to potential multicollinearity. I employ alternative variables or omit some independent variables from the regression and in case the values of the coefficients of the regressors remaining in the regression change considerably, this would support the assumption of multicollinearity. If the coefficients are fairly stable, it indicates little or no multicollinearity.

6.3.2. Endogeneity

The major challenge concerning the validity of my results is endogeneity. Endogeneity describes the phenomenon when independent variables are correlated with the error term causing the estimation of an affected variable via OLS to be inconsistent and biased. Endogeneity is primarily caused by measurement errors, simultaneity and omitted variable bias. The assumption of the existence or absence of endogeneity is based on a combination of intuition and knowledge from previous studies. Since I am only the second, after (Mares et al., 2018), to carry out an empirical study of wealth inequality at the country level, I have little to no guidance as to which variables are relevant for wealth inequality and which might be subject to endogeneity.

6.3.2.1. Sources of endogeneity

All the data I utilise is obtained from trustworthy sources, but the possibility of measurement errors cannot be ruled out. In fact, it is quite likely that some data series and some countries are subject to errors in data collection and that there are discrepancies in the comparability of data between countries.

Simultaneity poses a further issue, which is particularly challenging to address in my thesis. When simultaneity is present, the affected explanatory variable and the dependent variable have a reciprocal effect on each other, i.e. not only does the explanatory variable influence the dependent variable but also the dependent variable affects the explanatory variable.

Theoretically, simultaneity could be present for several variables in my findings. For instance, if the number of very wealthy people in a country is higher, i.e. if there is greater inequality of wealth, the wealthy may foster the development of financial markets because they seek to invest their capital. In turn more developed financial markets may reinforce wealth inequality since the wealthy in particular benefit from them. Similar rationales can be developed for, in principle, all my independent variables. However, the effect of wealth inequality on my independent variables could be somewhat smaller, because it is a relative variable not an absolute variable. To follow the previous example, a high wealth inequality in a poor country is less likely to contribute to the development of the financial markets than perhaps a less high inequality in a rich country, where more people overall have capital to invest.

The omitted variable bias occurs when a relevant explanatory variable is not included in the regression and is correlated with an included independent variable. Then the influence of the omitted variable on the dependent variable is captured in the included correlated independent variable and biases its estimates. In a panel data set, the omitted variables can be classified into two categories: time varying and time invariant variables.

6.3.2.2. Approaches to deal with endogeneity

The conventional approach to address endogeneity is to utilise instrument variables that are correlated with the endogenous variable but not with the error term. Suitable instruments can be either external instruments, i.e. other indicators correlated with the endogenous variable, or internal instruments, such as lags of the endogenous variable as in the Arellano-Bond model. However, identifying appropriate instrument variables can be immensely difficult and is based on many theoretical assumptions such as whether the instruments are indeed exogenous. The objective of my thesis is to provide an initial empirical insight into which variables are correlated with wealth inequality, since there is currently very limited understanding of this topic. Verifying my results on the scale necessary to provide complete certainty against endogeneity by finding and reasoning for appropriate instruments is beyond the scope of my research. Nonetheless, I perform a variety of robustness checks in order to achieve a certain degree of consistency of my results with respect to potential existing endogeneity.

Since I am not conducting an experimental study, but a study based on real world data, I cannot control for the quality of my data but instead have to rely to some extend on the quality of the available data. One option to verify my results against potential measurement error is to employ only a sub-sample of my countries, in which I can assume that the data quality is better, such as developed countries or the countries for which the wealth data was obtained from direct data[14]. I will conduct the corresponding robustness checks in the following.

As discussed above, simultaneity could be present for each of my independent variables. One typical way to assess robustness would be to employ the dynamic Arellano-Bond model. In this model, which is commonly based on first differences, higher lags of the assumed endogenous variable are used as instruments under the premise that they are exogenous. For this reason, I have decided not to apply the model in my thesis. The relationship between wealth inequality and my independent variables over time is not explored and since wealth inequality changes very slowly within a country over time, the higher lags of an endogenous variable could equally be endogenous just as well as my employed contemporaneous variables could be exogenous. In summary, insufficient knowledge about the determinants of wealth inequality is available, and my objective is to generate initial insights primarily about the more immediate effects. An extensive discussion and the overcoming of simultaneity can only be achieved through a multitude of studies, not by my work alone.

The most controllable source of endogeneity is the omitted variable bias. By the application of the fixed effects model the omitted variable bias of time invariant variables can be corrected. The fixed effects model achieves this by neutralising the unobserved country-specific effects via country-specific demeaning. Accordingly, my results are robust against time invariant omitted variable bias.

[14] As I described in chapter 3 'Data', Davies et al. (2017) compile their wealth data by using direct data mainly derived from household balance sheet data and surveys and for the countries which do not provide direct data they estimate the data. Arguably, the countries whose data is obtained from household balance sheets are less prone to measurement error than the estimated data.

Furthermore, I employ methods for time variant omitted variables bias to achieve a reasonable robustness for this cause of endogeneity. To begin with, lasso should reduce the likelihood of a time varying omitted variable bias, because in my study lasso should have selected the most meaningful variables out of a pool of 82 possible variables. Accordingly, at least the 73 indicators which were not selected by lasso should not be the cause of an omitted variable bias, since if they were more important than the nine selected by lasso, they should have been selected instead. Since the pool of 82 variables covers a wide range of economic, political and social indicators, it is reasonable to assume that many of the relevant determinants of wealth inequality are included.

In addition to this I apply several different methods to further validate the robustness of the variables selected by lasso. The first approach consists of using a variety of alternative variables for both my independent and my dependent variables. If regressing several variations of measurements of a parameter produce similar results, it indicates that the relationship between the parameter and wealth inequality is not accidental. In a sense, this constitutes an instrument variable estimation.

Secondly, I employ additional control variables on top of the seven baseline variables. If the coefficient and significance of a baseline variable changes substantially with the use of more control variables, this suggests that endogeneity constitutes an issue for that variable.

A specific case of omitted variable bias is the so-called spurious regression. Here, the omitted variable is a simple time trend. A spurious regression occurs for variables which are not stationary but also not co-integrated. The occurrence of a spurious regression is rather unusual in most fixed effects estimations because the time dimension of the panel data is typically too short. In my case, with T=19 the time dimension is in the intermediate range between short and long. In the detailed descriptive analysis of the Gini coefficient for wealth inequality that I conducted, no clear time trend was apparent. The Levin-Lin-Chu test for unit roots in panel data (Levin et al., 2002) rejects the null hypothesis that my dependent variable Gini wealth has a unit root. Consequently, the likelihood of the occurrence of a spurious regression should be low.

Nevertheless, in order to completely eliminate the possibility of a spurious regression, I utilize two techniques. First, I construct shortened panels in which I reduce the time dimension by taking averages of the values of my dependent and independent variables over several years. The probability of a spurious regression should decrease with a shorter time dimension.

Second, I employ two-way-fixed effects, i.e. fixed effects with year dummies. These time dummies should absorb a possible time trend and thus result in an alteration of the coefficient value and significance level of a spurious regression variable.

6.4. Robustness checks

In the following I will present the procedure and results of my robustness checks. Hereby, I describe the fixed effects estimation using only a sample of developed countries in more detail.

6.4.1. Developed Countries Sample

A robustness check which I analyse in greater detail is a fixed effects estimation based on a sample of developed countries and my baseline variables. Of the 129 countries in the baseline sample, only 31 are classified as advanced economies by the International Monetary Fund (2020a) and are used for the analysis.

As mentioned above, this reduction has the advantage that the developed countries have more reliable and comparable data. Therefore, employing only the developed countries should reduce possible endogeneity caused by measurement errors. Moreover, it should provide further insight into how the independent variables affect wealth inequality in developed countries, which also constitute a more homogeneous sample than in my baseline regression. However, these advantages come at a relatively high cost, namely a loss of information due to not utilizing more than three quarters of my available data. Furthermore, in a smaller sample the influence of idiosyncratic errors or outliers on the results increase.

Fixed effects is the efficient and consistent model according to the Breusch and Pagan Lagrangian multiplier test and the Hausmann test. Examining the collinearity between the parameters post fixed effects estimation indicates that no significant multicollinearity should be present. The results of fixed effects with heteroscedastic errors are displayed in Table 5.

Four of the seven variables are significant with only business regulations at the 1% significance level, education and labour force participation rate at the 5% level and control of corruption at the 10% level. Domestic credit again remains insignificant, but so are exports of goods and services and financial market depth in this estimation. Generally, the results here differ considerably from the baseline regression, since even the coefficients of the significant variables have changed by about 100% on average.

However, the significant coefficients exhibit a stronger effect on wealth inequality than in my baseline regression. Thus, the signs and direction of all coefficients are the same as in my baseline regression, but the magnitude of the value of the coefficients has increased. A labour force participation rate that is one percentage point higher affects wealth inequality by -0.4 points ceteris paribus instead of -0.232 points as in my baseline regression. An index point higher level in education is associated with a 0.18 higher wealth Gini instead of 0.068 in the baseline regression, for business regulations with 0.14 instead of 0.068 and for control of corruption with -0.169 instead of -0.069.

With regard to these four variables, the result can be interpreted as that although not necessarily the magnitude of the influence, the significance and direction seem robust. Thus, although the effects of the independent variables appear to be somewhat different in developed countries than for the whole country sample, I find supportive results for my baseline findings.

Table 5: Regression results developed countries sample

	Dependent variable:
	Wealth Gini
Exports/GDP	0.012
	(0.025)
Labour force participation	-0.400^{**}
	(0.185)
Financial market depth	-0.009
	(0.026)
Domestic credit	-0.006
	(0.014)
Education	0.180^{**}
	(0.087)
Business regulations	0.139^{***}
	(0.033)
Control of corruption	-0.168^{**}
	(0.082)
Observations	492
Within R^2	0.280
F Statistic	5.56^{***} (df = 7; 30)
Note:	*p<0.1,**p<0.05,***p<0.01

6.4.2. Further robustness checks

In the following I will briefly discuss a number of other robustness checks I have performed and highlight their main implications. In general, with the exception of one random effects estimation[15], they are all fixed effects regressions, as suggested respectively by the Breusch and Pagan Lagrangian multiplier test and the Hausmann test to be the consistent and efficient model respectively. Moreover, for all subsequent regressions I checked for multicollinearity which should not pose a problem and I employ heteroskedastic robust errors.

Detailed results for the following robustness checks are listed in the appendix in Table A5 to Table A12. Subsequently I provide the respective tables for each model specifications in parentheses.

Besides the robustness check with exclusively developed countries, I carry out two further checks in a similar manner. Once I consider all 172 countries for which I have data on wealth inequality, i.e. including small countries and countries with highly unbalanced data, to verify whether I obtain similar results by incorporating the maximum available information. Secondly, I only examine the 24 countries whose wealth data since the year 2000 is based on direct data obtained from household balance sheets or surveys[16] rather than being derived from regressions as I have discussed in chapter 3 'Data'. Therefore, I investigate whether I arrive at comparable results using only the most consistent wealth data.

My estimation with 172 countries yields essentially the same results as my baseline regression. Similarly, the results of the regression based on the countries with direct data[17] are almost identical to those of the developed countries estimation, which is not particularly surprising since 19 out of the 24 countries[18] with direct data are also included in the sample of developed countries. In summary, these two regressions support my previous findings (find the detailed regression outputs in Table A5 in the appendix).

Furthermore, I carry out a multitude of robustness checks on my baseline sample in which I replace independent variables and my dependent variable with alternative variables which

---

[15] For my sample of only countries with direct data.

16 I strictly use only those countries for which direct data are available for the entire period from 2000 to 2018. I refer to the origin of the data provided in the Global Wealth Report databooks as in Shorrocks et al. (2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018)). In chapter 3 'Data', I described that Davies et al. (2017) have direct data for 33 countries, however, this also includes nine countries which do not provide direct data for the entire period, which I consequently exclude here.

[17] According to the Hausmann test, the efficient and consistent model for this sample is random effects.

[18] The five countries that report direct data but are not developed countries are China, Hungary, India, Indonesia and South Africa. Since China and India are present, the direct data regression should have quite a large representational power in terms of covering a significant part of the world population.

represent a different measurement of the same indicator. Should the effect of an indicator be affirmed when using various measurements, this would support the robustness of its relevance. Of these robustness checks, the most natural one is the substitution of the variables exports of goods and services to GDP and the education index with their respective alternative variables which were also selected by lasso, the economic dimension of the Globalisation Index (Gygli et al., 2019) and the human capital index (Feenstra et al., 2015). In essence, the results of the estimation with the two alternative variables confirm the results of the baseline specification (see Table A6 in the appendix).

In addition, I replace the other independent variables with comparable parameters. I find that the overall value of the economic freedom index, which serves as a substitute for business regulations, and income inequality, which substitutes labour force participation rate[19], are likewise significant and confirm the respective linkage of the replaced variables: entrepreneurship is positively correlated with wealth inequality while labour income is crucial for the lower parts of the wealth distribution and positive shocks (like less unemployment or less income inequality) therein contribute to less inequality.

In contrast, I cannot establish further support for the variables domestic credit to the private sector, financial market depth[20] and control of corruption since their respective alternative variables are not significant. However, it is important to bear in mind that all the alternatives are drawn from the pool of my original 82 variables used by lasso, therefore they cannot be expected to have a strong significant impact on wealth inequality, otherwise lasso should have selected them (regression results can be found in Table A7 in the appendix).

Additionally, I employ the share of wealth of the top 1% of the richest as an alternative dependent variable once and the share of the top 10% another time on which I regress my baseline independent variables. Both series are also provided by the Credit Suisse Wealth Report databooks (Shorrocks et al., 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018).

[19] Clearly, income inequality and labour force participation rate measure different things, but no matching alternative variable for labour force participation rate is available to me. Utilising income inequality is motivated by the assumption that both indicators have an impact on earnings from labour, which should be crucial for the building or ownership of wealth for the bottom segments of the wealth distribution.

[20] Since the data for the alternative variables for financial market depth, such as stock market capitalisation or stock market total value traded, are not available for a sufficient number of countries in my baseline sample, I perform this robustness check on my developed country sample. It is important to note, however, that financial market depth in this sample was insignificant, so the expectations that the alternative variables will be significant are rather low (see the results in Table A8 in the appendix).

Since they represent only a part of the wealth distribution both variables provide imperfect measures of wealth inequality. However, they are the most suitable instruments that are available, and as they exhibit a correlation of 0.957 for the top 10% and 0.819 for the top 1% with the wealth Gini, I consider them to be quite representative. By and large, both regressions confirm my baseline findings, except that the depth of the financial markets variable has become insignificant (see Table A5 in the appendix).

Due to the above-mentioned limitation of using variables from the pool of lasso variables and that the wealth share of the top 1% or 10% are not perfect substitutes, these results should be interpreted as additional robustness for the variables that were repeatedly found to be significant - exports of goods and services or globalisation, the laboratory force participation rate or income from labour, education and business regulations or entrepreneurship - rather than negating the relevance of the variables that were not validated repeatedly - control of corruption and financial market depth.

Additionally, I perform robustness checks in which I incorporate additional control variables besides my regular variables. The control variables consist of general economic or social variables, such as GDP per capita or population size, or variables that have been found in the literature to influence wealth inequality, such as income inequality[21]. With the exception of my indicator for education, whose coefficient value fluctuates relatively strongly, both the significance and coefficient values of my baseline variables remain practically unchanged. This suggests that at least none of the control variables causes an omitted variable bias, except for maybe in education (see Table A9 in the appendix).

In order to better account for possible time variant omitted variable bias I employ two-way fixed effects for my baseline regression, i.e. fixed effects including year dummy variables. The annual effects appear not to be of great importance, as only two year dummies out of 18[22] are significant at the 10% level. In addition, six of my seven variables are virtually unchanged. On the other hand, education has become insignificant, and exhibits very high correlations with the year dummies in the post fixed effects correlation estimates. This suggests that education could be subject to spurious regression, i.e. that it only represents a time trend, which is absorbed here by the year dummies (see Table A10 in the appendix).

[21] Some of the control variables were estimated to be significant, but as I cannot evaluate their robustness, I will refrain from discussing them in detail.
[22] I have 19 year observations originally but through the fixed effects demeaning one year observation is diminished.

Furthermore, I carry out two regressions on condensed panels. Thereby I compute averages for certain intervals of years, whereby the intervals are all of the same length. Specifically, I calculate averages once over three and once over six years in the observational period from 2001 to 2018, i.e. with T=18[23], and accordingly obtain panels with a time dimension of six with three-year averages and once a time dimension of three with six-year averages. This enables me to assess whether the effects of my independent variables are persistent when averaged over a longer period of time. Furthermore, by contracting the time dimension I reduce the likelihood of a spurious regression.

In both specifications, four of the seven variables are significant, exports of goods and services, labour force participation rate, business regulations and control of corruption. For labour force participation rate, the value of the coefficient remains almost unchanged compared to the baseline regression. For the other three significant variables the coefficients are quite similar to my baseline results but are increasing in magnitude the longer the intervals of years for which the averages were calculated are. Financial market depth and education are insignificant, which may indicate insignificant average effects over a longer period or a spurious regression (see Table A5 in the appendix).

The results of lasso also confirm the respective effects of my independent variables on wealth inequality. The values of the coefficients in lasso are somewhat smaller in magnitude because they are subject to the shrinkage factor (see Table A3 and Table A4 in the appendix).

The education variable appears to be problematic because of all my variables, its coefficient value and the level of significance varies most from one regression to another. I suspect that the estimated relationship between the education index and wealth inequality is subject to a spurious regression, because in the model specifications with a stronger control for a time trend, as in two-way fixed effects and the shortened panels, education is always insignificant.

To get more clarity, I conduct a more detailed analysis of the development of education over time and find that only for 15% of the observations over years and countries the education indicator decreases and for 85% of the observations it increases. From the beginning of my sample period in 2000 to the end in 2018, the education index rose in 126 countries while it

[23] Since in my original sample with T=19 a prime number occurs as the length of the time dimension, I could not compute averages for equally long year intervals for it. Therefore, I omitted the oldest time observation (the year 2000).

fell in only 3. Furthermore, the Im–Pesaran–Shin test for unit roots in panel data (Im et al., 2003) indicates that unit roots are present in education[24].

Those are strong indications that the estimated correlation between education and inequality stems from that education captures the positive part of the development of the wealth Gini over time, in other words it is subject to a spurious regression.

As a consequence, I regard education as a non-significant variable. To verify the results of my other variables, I perform regressions both without the education indicator and with an education measure[25] that is not subject to such a dynamic and find that my remaining independent variables remain quite stable and thus should be valid[26] (see Table A11 in the appendix).

Based on this approach, I perform a final type of robustness check in which I omit one of the independent variables from the regression at a time. The significance levels and the value of the coefficients of the independent variables remaining in the regression remain largely stable, with the exception that when the business regulations index is omitted, control of corruption becomes insignificant. Hence, I infer that my estimates are quite reliable, particularly with respect to multicollinearity (see Table A12 in the appendix).

In summary, I find that three of my seven variables - labour force participation rate, business regulations and control of corruption - have a very high degree of robustness as they are significant in virtually all model specifications and are quite stable in the size of their coefficients. Exports of goods and services to GDP also displays strong robustness, since it is significant in all robustness checks but in the smaller samples and its alternative variables are also significant and positively correlated with wealth inequality.

Financial market depth exhibits moderate robustness, being significant in many robustness checks, but also being insignificant in others. Education and domestic credit to private sector can be assessed as insignificant, the former because of a probable spurious regression and the latter because it is consistently estimated as insignificant.

[24] I cannot employ the Levin–Lin–Chu test here as I do for the Gini coefficient of wealth previously since education is unbalanced and the Levin–Lin–Chu test does not allow for unbalanced data. For the wealth Gini the Im–Pesaran–Shin test and the Levin–Lin–Chu test yield the same result, that no unit root should be present.

[25] Adjusted savings education expenditure (in % of GDP)

[26] To be more precise, I find that the magnitude of the coefficients increases slightly, but it can be an advantage for validity when a variable like the education index filters out a trend, so I rely on my previous estimates.

The fact that domestic credit to the private sector was selected by lasso but is not significant in any of my subsequent model specifications can be explained by the different procedures of lasso and fixed effects, as described in chapter 4.3 'Methodology'.

Another contributing element could be that domestic credit was selected by lasso with homoscedastic errors, whereas I consistently chose heteroscedastic errors in my fixed effects estimates in order to ensure heteroscedasticity robust results. In general, however, the selection of lasso has proven to be successful, since only one of the selected variables consequently is estimated insignificant.

Regarding endogeneity, the findings generally suggest that there is no major problem beyond that associated with the education index, since my results demonstrate a relatively high robustness across different samples, model specifications, alternative variables and with the inclusion of control variables. Although I cannot rule out the possibility of the occurrence of endogeneity entirely, with respect to the limitations of my work discussed in the chapter 6.3.2 'Endogeneity', I can conclude that endogeneity should be largely controlled for.

7. Discussion

7.1. General discussion of my results

Both main aspects of my approach, the identification of influential variables and the subsequent quantification of their influence, are quite successful. Through lasso, I select nine variables from a large pool of possible variables. Out of these nine variables, I employ seven which describe distinct economic and societal aspects, two of which relate directly to finance.

For five of the seven variables - exports of goods and services to GDP, labour force participation rate, financial market depth, business regulations, control of corruption - a fairly robust relationship with wealth inequality can be observed which is consistent with the literature and theory on wealth inequality as well as with economic intuition.

Thereby, my perhaps most unexpected finding is that the results do not indicate that financial development is as important for wealth inequality as assumed by the theory and the literature, since lasso selects only two of many possible financial variables, and those selected exhibit only a modest influence on inequality. A possible explanation for the moderate effect of finance could be that by and large the different segments of the wealth distribution benefit approximately equally from financial development leaving the degree of inequality largely unchanged. Further potential explanations are that the temporal relationship between finance and wealth inequality is longer than has been modelled by me, or that finance only exerts a strong influence in combination with other variables.

More specifically, for the two financial variables included in my baseline regression, I find that a greater depth of financial markets is associated with increased inequality while domestic credit to the private sector, a measure of the depth of financial institutions, appears not to be a significant factor. The overall effect of finance therefore appears to be modestly inequality increasing.

The effect of financial market depth in increasing inequality can be attributed to that wealthy people are the primary participants in the financial markets and thus benefit disproportionately from a greater depth of the financial markets (Mares et al., 2018; Roine and Waldenström, 2015; Saez and Zucman, 2014). In fact, the only other empirical study, by Mares et al. (2018), precisely identified this indicator as being related to higher wealth inequality.

Domestic credit to the private sector's absent impact may be due to that the different segments of the wealth distribution benefit roughly evenly from credits granted and thus tend not to overly favour a portion of the population which would drive inequality in one direction or the other. For instance, loans for house purchase would tend to benefit the middle bracket of the wealth distribution, thereby reducing wealth inequality, while credits for businesses are more likely to benefit the upper bracket, leading to an increase in inequality (Bozio et al., 2017).

For the non-finance related variables, I find relationships to wealth inequality that correspond to the theory and literature. The positive correlation between globalisation, proxied as exports of goods and services to GDP, and wealth inequality could be explained by the fact that globalisation's greatest beneficiaries are the wealthy and highest qualified, because they are able to capitalise most on the opportunities of a more globalized world (Dabla-Norris et al., 2015; Roine and Waldenström, 2015). Similarly to financial market depth, Mares et al. (2018) found this exact variable as having the same effect as in my study by being positively correlated to wealth inequality.

Also linked to higher wealth inequality is a more beneficiary regulatory environment for starting and running a business which constitutes an obvious connection, as entrepreneurship is the most important factor determining the possibility of generating large lifetime earnings and therefore wealth (Meh, 2005; Nardi and Fella, 2017; Pástor and Veronesi, 2016). However, the relationship may also be caused by countries displaying different weighting of the importance of entrepreneurship and social equity, and countries that promote more entrepreneurship generally create a more business-friendly environment, including lower social costs and minimum wages than countries that aim for more equality.

Particularly plausible are the reducing effects on wealth inequality of the labour force participation rate, since for the lower part of the wealth distribution labour income constitutes the single most relevant mean to acquire wealth. Ergo the least wealthy benefit disproportionately from reduced unemployment which consequently decreases inequality (Bozio et al., 2017; Nardi et al., 2016; Scholz and Seshadri, 2007).

Furthermore, corruption tends to facilitate the concentration of power and wealth in the hands of the few who already possess them (Alesina and Angeletos, 2005; Bagchi and Svejnar, 2015). Which is in accordance with my findings that a higher control of corruption implies lower wealth inequality.

Since I consider the education variable, the education index, subject to a spurious regression and its alternative variable, education expenditure, is insignificant, I do not observe a significant effect of education on wealth inequality. This could be due to several reasons, one theoretically likely possibility is that the relationship over time between education and inequality differs from the rather immediate one I modelled. One might assume that the effect of improved education on wealth inequality is considerably lagged, since the primary beneficiaries are children or young people still in education. Several years would have to pass before they enter employment and are able to accumulate wealth and thus influence the inequality level.

It is worth mentioning that the size of the coefficients in my results are all rather small. For example in my baseline regression, a variation of one index point in any of the index variables[27] only results in a maximum change of 0.069 Gini points. Consequently, in a fixed effects model, the index would have to be at least 14.5 points higher than its average value to yield ceteris paribus an average inequality that is one Gini point higher. Even the coefficient of labour force participation, which has the largest absolute value of -0.232 in my baseline regression, implies that unemployment in a given country and year would have to be more than four percentage points lower than its average to translate into a reduction of one Gini point in inequality. In a real world economy, this is a substantial figure.

[27] Financial market depth, business regulations, control of corruption

The reason for the small size of the coefficients can be attributed mainly to the use of fixed effects, which primarily accounts for the within variance, which, as I pointed out in the descriptive analysis of wealth inequality, is relatively low compared to the between variance. Additionally, wealth inequality tends to change rather gradually within countries. Furthermore, with my model specifications, I investigate primarily the short-term effects of my regressors on my dependent variable, while long-term effects presumably exist as well.

7.2. Implications of wealth inequality

For two major reasons, I do not make any direct policy recommendations based on my findings. The first one is that only few studies on the effects of wealth inequality exist due to the same reasons[28] that there is little literature on the determinants of wealth inequality. The available studies suggest that wealth inequality may reduce economic growth (Bagchi and Svejnar, 2015; Islam and McGillivray, 2020), but also may stimulate entrepreneurship (Frid et al., 2016), lead to higher per capita CO2 emissions in high-income countries (Knight et al., 2017), and negatively affect health in low-income countries, especially through children's under-nutrition (Hong et al., 2006; Hong and Mishra, 2006).

The second reason I abstain from policy recommendations is based on the fact that my dependent variable, the Gini coefficient for wealth, is a relative variable. Hence, it does not measure the absolute level of wealth but its distribution. Most of my independent variables are likely to affect the lives and wealth of the entire population of a country, and it is probable that the effect of the variable on the level of wealth is equally increasing (or decreasing) for each segment of society. However, if the decreasing or increasing effect is disproportionately large for a particular part of the wealth distribution, inequality will shrink or rise depending on which part of the population is more affected.

Ergo without considering the absolute effects, the impact of policies aimed at reducing inequality could lead to all segments of the population losing wealth, merely by different magnitudes. In order to assess the full implications of my explanatory variables on wealth, both in absolute terms and in distribution, a study would be necessary that examines the effects of my variables on the absolute level or, at best, includes both relative and absolute aspects.

In summary, without knowledge about the absolute level and distribution of wealth alike, or reliable understanding provided by other studies on the effects of inequality alone, a complete picture of the consequences of wealth inequality cannot be derived. Hence the discussion of

[28] primarily the lack of available data

whether the positive aspects of inequality or the negative aspects predominate can only be answered based on believes in a normative manner from which I abstain here.

7.3. Limitations and outlook

My thesis provides an initial insight into the determinants of wealth inequality, and therefore cannot conclusively answer all the relevant questions. The limitations of my study present scope for future research, and I briefly mention the most noteworthy challenges.

First of all, the relationship between my independent variables and wealth inequality is modelled primarily as short-term. Longer-term effects of the explanatory variables may also be expected, especially since wealth inequality is a rather slowly evolving variable. For this purpose, employing explanatory variables in lags of varying lengths can provide an approach to gain further insight.

Beyond that, a dynamic development of wealth inequality is also possible, whereby an already high level of inequality leads to even higher inequality in the future or vice versa. This could occur through a reinforcing circle in conjunction with independent variables, i.e. a form of simultaneity, or directly. Hereby, dynamic panel data models represent an extension of my approach to investigate the issues described more deeply.

Another approach is to include combined effects of independent variables in the model through the use of interaction terms, which could be especially interesting regarding finance since in the theory often saving in combination with other variables leads to inequality.

In summary, since so far relatively little is known concerning the determinants of wealth inequality, the true functional form of the relationship is uncertain. Confidence in this regard will be primarily achieved through the application of different model specifications and methods, which can also extend beyond those of the panel data models.

7.4. Discussion summary

My thesis provides new insight into the determinants of wealth inequality by identifying several relevant variables. Since little empirical research has been conducted so far, I expand and can partly confirm previous findings on the pool of influential variables and thus strengthen the empirical basis for the theory on wealth inequality. Moreover, my results can serve as a basis for subsequent studies to extend beyond the first step of identifying variables and investigate in more detail the functional relationship between the independent variables and wealth inequality.

8. Conclusion

Based on relatively new wealth panel data from the Credit Suisse Wealth Reports, I was able to identify and quantify the influence of five variables that assert a significant effect on wealth inequality. I find that greater globalisation, a more business-friendly regulatory environment and greater depth of financial markets are associated with higher wealth inequality, while a higher labour force participation rate and stronger control of corruption are associated with lower inequality.

Labour force participation rate, business regulations and control of corruption in particular show high robustness across different model specifications, as well as globalisation to a slightly lower degree, while for financial market depth I find a moderate robustness. Since I identify only one significant financial variable that is moderately robust, I observe a moderate to weak relationship between financial development and wealth inequality. Since financial development serves as an approximation for saving in my study, I do not find empirical support for the emphasis on the importance of saving for wealth inequality which it is given in the theory. Instead non-financial variables appear to be more relevant for wealth inequality than financial ones.

My thesis contributes to the hitherto scarce empirical literature on wealth inequality. The five determinants which I identified and their implications for wealth inequality are consistent with what has been discussed in the theory. Moreover, the only other cross-country study of which I am aware (Mares et al., 2018), identified two variables which are identical to my variables for globalisation and financial market depth. Mares et al. (2018) likewise observed an inequality-increasing effect of these variables.

Overall, my results reflect the existing literature and theory on the determinants of wealth inequality, with the exception that I do not find evidence for the major importance of financial development.

Through my study, I provide additional empirical evidence for both the academic and the public debate on wealth inequality. In case that the trend of increasing wealth inequality, which has been pointed out by many authors (Piketty and Zucman, 2014; Roine and Waldenström, 2015) and which has seemingly accelerated by the Corona virus, continues, the discussion about it will intensify accordingly. In this context, science must present unbiased and data-based insights into both the causes and the consequences of wealth inequality. Covid-19 has not only highlighted the extent of wealth inequality but also, in general, the important role of science in the effective and sensible handling of complex problems.

References

Ahrens, A., Hansen, C. B. and Schaffer, M. E. (2020) 'lassopack: Model selection and prediction with regularized regression in Stata', *The Stata Journal: Promoting communications on statistics and Stata*, vol. 20, no. 1, pp. 176–235. DOI: 10.1177/1536867X20909697.

Alesina, A. and Angeletos, G.-M. (2005) 'Corruption, inequality, and fairness', *Journal of Monetary Economics*, vol. 52, no. 7, pp. 1227–1244. DOI: 10.1016/j.jmoneco.2005.05.003.

Alvaredo, F., Atkinson, A. B., Piketty, T. and Saez, E. (2013) 'The Top 1 Percent in International and Historical Perspective', *Journal of Economic Perspectives*, vol. 27, no. 3, pp. 3–20.

Bach, L., Calvet, Laurent, E. and Sodini, P. (2015) 'Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy', *Working Papers hal-02002692*, HAL.

Bagchi, S. and Svejnar, J. (2015) 'Does wealth inequality matter for growth? The effect of billionaire wealth, income distribution, and poverty', *Journal of Comparative Economics*, vol. 43, no. 3, pp. 505–530.

Belloni, A., Chernozhukov, V. and Chen, D., Hansen, C. (2012) 'Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain', *Econometrica*, vol. 80, no. 6, pp. 2369–2429.

Belloni, A., Chernozhukov, V., Hansen, C. and Kozbur, D. (2016) 'Inference in High-Dimensional Panel Models With an Application to Gun Control', *Journal of Business & Economic Statistics*, vol. 34, no. 4, pp. 590–605.

Belloni, A., Chernozhukov, V. and Wang, L. (2011) 'Square-root lasso: pivotal recovery of sparse signals via conic programming', *Biometrika*, vol. 98, no. 4, pp. 791–806.

Benhabib, J., Bisin, A. and Luo, M. (2017) 'Earnings Inequality and Other Determinants of Wealth Inequality', *American Economic Review*, vol. 107, no. 5, pp. 593–597.

Benhabib, J., Bisin, A. and Zhu, S. (2011) 'The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents', *Econometrica*, vol. 79, no. 1, pp. 123–157.

Benhabib, J., Bisin, A. and Zhu, S. (2015) 'The wealth distribution in Bewley economies with capital income risk', *Journal of Economic Theory*, vol. 159, pp. 489–515.

Bernheim, B. D., Skinner, J. and Weinberg, S. (2001) 'What Accounts for the Variation in Retirement Wealth Among U.S. Households?', *American Economic Review*, vol. 91, no. 4, pp. 832–857.

Bittencourt, M., Chang, S., Gupta, R. and Miller, S. M. (2019) 'Does financial development affect income inequality in the U.S. States?', *Journal of Policy Modeling*, vol. 41, no. 6, pp. 1043–1056.

Blandin, A. and Peterman, W. B. (2019) 'Taxing capital? The importance of how human capital is accumulated', *European Economic Review*, vol. 119, pp. 482–508.

Bloomberg L.P. (2020) *The Bloomberg Billionaires Index* [Online], New York. Available at https://www.bloomberg.com/billionaires/ (Accessed 31 December 2020).

Blume, L. E. and Durlauf, S. N. (2015) 'Capital in the Twenty-First Century: A Review Essay', *Journal of Political Economy*, vol. 123, no. 4, pp. 749–777.

Bozio, A., Emmerson, C., O'Dea, C. and Tetlow, G. (2017) 'Do the rich save more? Evidence from linked survey and administrative data', *Oxford Economic Papers*, vol. 61, p. 739.

Cagetti, M. and Nardi, M. de (2008) 'Wealth Inequality: Data and Models', *Macroeconomic Dynamics*, vol. 12, S2, pp. 285–313.

Chinn, M. D. and Ito, H. (2008) 'A New Measure of Financial Openness', *Journal of Comparative Policy Analysis: Research and Practice*, vol. 10, no. 3, pp. 309–322. DOI: 10.1080/13876980802231123.

Credit Suisse Research Institute (2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018) *Global Wealth Report,* Credit Suisse Research Institute [Online]. Available at https://www.credit-suisse.com/about-us/en/reports-research/global-wealth-report.html.

Dabla-Norris, E., Tsounta, E., Ricka, F., Kochhar, K. and Suphaphiphat, N. (2015) *Causes and Consequences of Income Inequality: A Global Perspective* [Online], Washington, D.C, International Monetary Fund. Available at http://www.elibrary.imf.org/view/IMF006/22594-9781513555188/22594-9781513555188/22594-9781513555188.xml.

Davies, J. B., Lluberas, R. and Shorrocks, A. F. (2017) 'Estimating the Level and Distribution of Global Wealth, 2000-2014', *Review of Income and Wealth*, vol. 63, no. 4, pp. 731–759.

Davies, J. B., Sandström, S., Shorrocks, A. and Wolff, E. N. (2011) 'The Level and Distribution of Global Household Wealth', *The Economic Journal*, vol. 121, no. 551, pp. 223–254.

Fagereng, A., Guiso, L., Malacrino, D. and Pistaferri, L. (2016) 'Heterogeneity in Returns to Wealth and the Measurement of Wealth Inequality', *American Economic Review*, vol. 106, no. 5, pp. 651–655.

Feenstra, R. C., Inklaar, R. and Timmer, M. P. (2015) 'The Next Generation of the Penn World Table: Penn World Table version 9.1', *American Economic Review*, 105 (10), pp. 3150–3182. DOI: 10.15141/S50T0R.

Flavin, M. and Yamashita, T. (2002) 'Owner-Occupied Housing and the Composition of the Household Portfolio', *American Economic Review*, vol. 92, no. 1, pp. 345–362.

Fraser Institute (2020a) *Business regulations index: Economic Freedom of the World* [Online]. Available at https://www.fraserinstitute.org/studies/economic-freedom (Accessed 17 December 2020).

Fraser Institute (2020b) *Economic Freedom Index* [Online]. Available at https://www.fraserinstitute.org/economic-freedom/dataset?geozone=world&year=2018&page=dataset&min-year=2&max-year=0&filter=0 (Accessed 21 December 2020).

Freedom House (2020) *Freedom in the World: Reports* [Online]. Available at https://freedomhouse.org/report/freedom-world/2020/leaderless-struggle-democracy (Accessed 21 December 2020).

Freixas, X. and Rochet, J.-C. (1997) *Microeconomics of banking*, Cambridge, Mass., London, MIT Press.

Frid, C. J., Wyman, D. M. and Coffey, B. (2016) 'Effects of wealth inequality on entrepreneurship', *Small Business Economics*, vol. 47, no. 4, pp. 895–920.

Giles, J. A. and Giles, D. E. A. (1993) 'Pre-test estimation and testing in econometrics: Recent developments', *Journal of Economic Surveys*, vol. 7, pp. 145–197 [Online]. Available at http://web.uvic.ca/~dgiles/blog/Giles%20and%20Giles.pdf.

Gleditsch, N. P., Wallensteen, P., Eriksson, M. and Sollenberg, Margareta, Strand, Håvard (2002) 'Armed Conflict 1946-2001: A New Dataset', *Journal of Peace Research*, 39(5) [Online]. Available at https://ucdp.uu.se/downloads/index.html.

Gygli, S., Haelg, F., Potrafke, N. and Sturm, J.-E. (2019) 'The KOF Globalisation Index – revisited', *The Review of International Organizations*, vol. 14, no. 3, pp. 543–574. DOI: 10.1007/s11558-019-09344-2.

Hendricks, L. (2007) 'Retirement Wealth and Lifetime Earnings', *International Economic Review*, vol. 48, no. 2, pp. 421–456.

Hong, R., Banta, J. E. and Betancourt, J. A. (2006) 'Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh', *International journal for equity in health*, vol. 5, p. 15.

Hong, R. and Mishra, V. (2006) 'Effect of Wealth Inequality on Chronic Under-Nutrition in Cambodian Children', *Journal of Health, Population and Nutrition*, vol. 24, no. 1, pp. 89–99 [Online]. Available at www.jstor.org/stable/23499271 (Accessed 26 December 2020).

Huggett, M., Ventura, G. and Yaron, A. (2011) 'Sources of Lifetime Inequality', *American Economic Review*, vol. 101, no. 7, pp. 2923–2954.

Im, K. S., Pesaran, M.H. and Shin, Y. (2003) 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics*, vol. 115, no. 1, pp. 53–74. DOI: 10.1016/s0304-4076(03)00092-7.

International Monetary Fund (2020a) *Financial Development Index* [Online], Washington, D.C. Available at https://data.imf.org/?sk=F8032E80-B36C-43B1-AC26-493C5B1CD33B (Accessed 23 December 2020).

International Monetary Fund (2020b) *Financial Markets Depth: Financial Development Index* [Online], Washington, D.C. Available at https://data.imf.org/?sk=F8032E80-B36C-43B1-AC26-493C5B1CD33B (Accessed 5 December 2020).

Islam, M. R. and McGillivray, M. (2020) 'Wealth inequality, governance and economic growth', *Economic Modelling*, vol. 88, pp. 1–13. DOI: 10.1016/j.econmod.2019.06.017.

Jauch, S. and Watzka, S. (2016) 'Financial development and income inequality: a panel data approach', *Empirical Economics*, vol. 51, no. 1, pp. 291–314.

Jones, C. I. (2015) ‘Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality’, *Journal of Economic Perspectives*, vol. 29, no. 1, pp. 29–46.

Kaymak, B. and Poschke, M. (2016) ‘The evolution of wealth inequality over half a century: The role of taxes, transfers and technology’, *Journal of Monetary Economics*, vol. 77, pp. 1–25.

Knight, K. W., Schor, J. B. and Jorgenson, A. K. (2017) ‘Wealth Inequality and Carbon Emissions in High-income Countries’, *Social Currents*, vol. 4, no. 5, pp. 403–412.

Kopczuk, W. (2015) ‘What Do We Know about the Evolution of Top Wealth Shares in the United States?’, *Journal of Economic Perspectives*, vol. 29, no. 1, pp. 47–66.

Levin, A., Lin, C.-F. and James Chu, C.-S. (2002) ‘Unit root tests in panel data: asymptotic and finite-sample properties’, *Journal of Econometrics*, vol. 108, no. 1, pp. 1–24. DOI: 10.1016/S0304-4076(01)00098-7.

Mankiw, N. G. (2015) ‘Yes, r > g. So What?’, *American Economic Review*, vol. 105, no. 5, pp. 43–47.

Mares, J., Horvath, R. and Hasan, I. (2018) ‘Finance and Wealth Inequality’*, IES FSV, Charles University, IES Working Papers 35*.

Meh, C. A. (2005) ‘Entrepreneurship, wealth inequality, and taxation’, *Review of Economic Dynamics*, vol. 8, no. 3, pp. 688–719.

Nardi, M. de and Fella, G. (2017) ‘Saving and wealth inequality’, *Review of Economic Dynamics*, vol. 26, pp. 280–300.

Nardi, M. de, Fella, G. and Pardo, G. P. (2016) ‘The Implications of Richer Earnings Dynamics for Consumption and Wealth: Working Paper 21917’, *nber Working Paper Series*.

Nardi, M. de and Yang, F. (2016) ‘Wealth inequality, family background, and estate taxation’, *Journal of Monetary Economics*, vol. 77, pp. 130–145.

Pástor, L. and Veronesi, P. (2016) ‘Income inequality and asset prices under redistributive taxation’, *Journal of Monetary Economics*, vol. 81, pp. 1–20.

Pettersson, T. and Öberg, M. ‘Organized violence, 1989-2019’, *Journal of Peace Research*, vol. 2020, 57(4) [Online]. Available at https://ucdp.uu.se/downloads/index.html.

Piketty, T. and Goldhammer, A. (2014) *Capital in the twenty-first century*, Cambridge Massachusetts, The Belknap Press of Harvard University Press.

Piketty, T. and Zucman, G. (2014) ‘Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010 *’, *The Quarterly Journal of Economics*, vol. 129, no. 3, pp. 1255–1310.

Powell, J. and Thyne, C. (2011) ‘Global Instances of Coups from 1950-Present’, *Journal of Peace Research*, 48(2), pp. 249–259 [Online]. Available at https://www.jonathanmpowell.com/coup-detat-dataset.html (Accessed 21 December 2020).

Roine, J. and Waldenström, D. (2009) ‘Wealth Concentration over the Path of Development: Sweden, 1873-2006’, *Scandinavian Journal of Economics*, vol. 111, no. 1, pp. 151–187.

Roine, J. and Waldenström, D. (2015) 'Long-Run Trends in the Distribution of Income and Wealth', in Atkinson, A. B. and Bourguignon, F. (eds) *Handbook of income distribution. Vols. 2A-2B,* Amsterdam, Elsevier/North Holland, pp. 469–592.

Saez, E. and Zucman, G. (2014) 'Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data'*, nber Working Paper Series*.

Scartascini, C., Cruz, C. and Keefer, P. (2018) *The Database of Political Institutions 2017 (DPI2017),* Inter-American Development Bank. Numbers for Development. [Online]. Available at https://mydata.iadb.org/Reform-Modernization-of-the-State/.

Scholz, J. K. and Seshadri, A. (2007) 'Children and Household Wealth', *SSRN Electronic Journal*.

Shorrocks, A., Davies, J. B. and Lluberas, R. (2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018) *Global Wealth Databook,* Credit Suisse Research Institute [Online]. Available at https://www.credit-suisse.com/about-us/en/reports-research/global-wealth-report.html.

Solt, F. (2020) 'Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database', *Social Science Quarterly*, vol. 101, no. 3, pp. 1183–1199.

Svirydzenka, K. (2016) *Introducing a new broad-based index of financial development*, Washington, D.C., International Monetary Fund.

The World Bank (2020a) *Control of Corruption: Estimate: Worldwide Governance Indicators Database* [Online]. Available at https://data.worldbank.org/indicator/FS.AST.PRVT.GD.ZS (Accessed 17 December 2020).

The World Bank (2020b) *Domestic credit to private sector (% of GDP): World Development Indicators Database* [Online]. Available at https://data.worldbank.org/indicator/FS.AST.PRVT.GD.ZS (Accessed 17 December 2020).

The World Bank (2020c) *Exports of goods and services (% of GDP): World Development Indicators Database* [Online]. Available at https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS (Accessed 17 December 2020).

The World Bank (2020d) *Global Financial Development Database* [Online]. Available at https://datacatalog.worldbank.org/dataset/global-financial-development (Accessed 21 December 2020).

The World Bank (2020e) *Labour force participation rate, total (% of total population ages 15+): World Development Indicators Database* [Online]. Available at https://data.worldbank.org/indicator/SL.TLF.CACT.ZS (Accessed 17 December 2020).

The World Bank (2020f) *World Development Indicators Database* [Online]. Available at https://databank.worldbank.org/source/world-development-indicators/preview/on (Accessed 21 December 2020).

The World Bank (2020g) *Worldwide Governance Indicators Database* [Online]. Available at https://datacatalog.worldbank.org/dataset/worldwide-governance-indicators (Accessed 21 December 2020).

The World Bank (2020h) *COVID-19 to Add as Many as 150 Million Extreme Poor by 2021* [Online], Washington, D.C. Available at https://www.worldbank.org/en/news/press-release/2020/10/07/covid-19-to-add-as-many-as-150-million-extreme-poor-by-2021 (Accessed 21 December 2020).

Tibshirani, R. (1996) ‘Regression Shrinkage and Selection Via the Lasso’, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288.

Toh, M. (2020) ‘Jeff Bezos is now worth a whopping $200 billion’, *CNN Business*, 28 August [Online]. Available at https://edition.cnn.com/2020/08/27/tech/jeff-bezos-net-worth-200-billion-intl-hnk/index.html (Accessed 21 December 2020).

United Nations (2020) *Education index: Human Development Reports* [Online]. Available at http://hdr.undp.org/en/indicators/103706 (Accessed 17 December 2020).

Wolff, E. N. and Gittleman, M. (2014) ‘Inheritances and the distribution of wealth or whatever happened to the great inheritance boom?’, *The Journal of Economic Inequality*, vol. 12, no. 4, pp. 439–468.

Appendix

Table A1: Countries by region and income groups

Region (IMF)	Country	Income group (IMF)
Africa	Angola	Emerging market
	Benin	Low-income developing country
	Burkina Faso	Low-income developing country
	Burundi	Low-income developing country
	Cameroon	Low-income developing country
	Central African Republic	Low-income developing country
	Chad	Low-income developing country
	Congo	Low-income developing country
	Congo (Democratic Republic of the)	Low-income developing country
	Ethiopia	Low-income developing country
	Ghana	Low-income developing country
	Guinea	Low-income developing country
	Kenya	Low-income developing country
	Liberia	Low-income developing country
	Madagascar	Low-income developing country
	Malawi	Low-income developing country
	Mali	Low-income developing country
	Mozambique	Low-income developing country
	Niger	Low-income developing country
	Nigeria	Emerging market
	Rwanda	Low-income developing country
	Senegal	Low-income developing country
	Sierra Leone	Low-income developing country
	South Africa	Emerging market
	Tanzania (United Republic of)	Low-income developing country
	Togo	Low-income developing country

	Uganda	Low-income developing country
	Zambia	Low-income developing country
	Zimbabwe	Low-income developing country
Asia and Pacific	Australia	Advanced market
	Bangladesh	Low-income developing country
	Cambodia	Low-income developing country
	China	Emerging market
	Hong Kong, China (SAR)	Advanced market
	India	Emerging market
	Indonesia	Emerging market
	Japan	Advanced market
	Korea (Republic of)	Advanced market
	Lao People's Democratic Republic	Low-income developing country
	Malaysia	Emerging market
	Mongolia	Emerging market
	Myanmar	Low-income developing country
	Nepal	Low-income developing country
	New Zealand	Advanced market
	Papua New Guinea	Low-income developing country
	Philippines	Emerging market
	Singapore	Advanced market
	Sri Lanka	Emerging market
	Thailand	Emerging market
	Viet Nam	Emerging market
Europe	Albania	Emerging market
	Austria	Advanced market
	Belarus	Emerging market
	Belgium	Advanced market
	Bosnia and Herzegovina	Emerging market

Bulgaria	Emerging market
Croatia	Emerging market
Czechia	Advanced market
Denmark	Advanced market
Estonia	Advanced market
Finland	Advanced market
France	Advanced market
Germany	Advanced market
Greece	Advanced market
Hungary	Emerging market
Ireland	Advanced market
Israel	Advanced market
Italy	Advanced market
Latvia	Advanced market
Lithuania	Advanced market
Moldova (Republic of)	Low-income developing country
Netherlands	Advanced market
Norway	Advanced market
Poland	Emerging market
Portugal	Advanced market
Romania	Emerging market
Russian Federation	Emerging market
Serbia	Emerging market
Slovakia	Advanced market
Slovenia	Advanced market
Spain	Advanced market
Sweden	Advanced market
Switzerland	Advanced market
Turkey	Emerging market
Ukraine	Emerging market

	United Kingdom	Advanced market
Middle East and Central Asia	Algeria	Emerging market
	Armenia	Emerging market
	Azerbaijan	Emerging market
	Egypt	Emerging market
	Georgia	Emerging market
	Iran (Islamic Republic of)	Emerging market
	Iraq	Emerging market
	Jordan	Emerging market
	Kazakhstan	Emerging market
	Kuwait	Emerging market
	Kyrgyzstan	Low-income developing country
	Lebanon	Emerging market
	Libya	Emerging market
	Mauritania	Low-income developing country
	Morocco	Emerging market
	Oman	Emerging market
	Pakistan	Emerging market
	Saudi Arabia	Emerging market
	Sudan	Low-income developing country
	Syrian Arab Republic	Emerging market
	Tajikistan	Low-income developing country
	Tunisia	Emerging market
	United Arab Emirates	Emerging market
	Yemen	Low-income developing country
Western Hemisphere	Argentina	Emerging market
	Bolivia (Plurinational State of)	Emerging market
	Brazil	Emerging market
	Canada	Advanced market
	Chile	Emerging market

	Colombia	Emerging market
	Costa Rica	Emerging market
	Ecuador	Emerging market
	El Salvador	Emerging market
	Haiti	Low-income developing country
	Jamaica	Emerging market
	Mexico	Emerging market
	Nicaragua	Low-income developing country
	Panama	Emerging market
	Paraguay	Emerging market
	Peru	Emerging market
	United States	Advanced market
	Uruguay	Emerging market
	Venezuela (Bolivarian Republic of)	Emerging market
Not in baseline regression included countries (due to small population size of country or low availability of data)	Afghanistan	
	Antigua and Barbuda	
	Aruba	
	Bahamas	
	Bahrain	
	Barbados	
	Belize	
	Botswana	
	Brunei Darussalam	
	Comoros	
	Cyprus	
	Djibouti	
	Dominica	
	Equatorial Guinea	
	Eritrea	

Fiji
Gabon
Gambia
Grenada
Guinea-Bissau
Guyana
Iceland
Lesotho
Luxembourg
Maldives
Malta
Mauritius
Montenegro
Namibia
Qatar
Saint Lucia
Saint Vincent and the Grenadines
Samoa
Sao Tome and Principe
Seychelles
Solomon Islands
Suriname
Taiwan, China
Timor-Leste
Tonga
Trinidad and Tobago
Turkmenistan
Vanuatu

Table A2: Pool of variables for lasso

Category	Subcategory	Variable	Source
Globalisation	***General***	Exports of goods and services (% of GDP)	(The World Bank, 2020c)
		Foreign direct investment, net inflows (% of GDP)	(The World Bank, 2020f)
		Globalisation index	(Gygli et al., 2019)
		Globalisation index economic dimension	(Gygli et al., 2019)
		Globalisation index finance dimension	(Gygli et al., 2019)
		Tariffs	(Fraser Institute, 2020b)
		Regulatory trade barriers	Fraser Institute (2020b)
		Capital controls	Fraser Institute (2020b)
		Freedom to trade internationally	Fraser Institute (2020b)
Finance	***General***	Financial development	(International Monetary Fund, 2020a)
		Financial institutions	(International Monetary Fund, 2020a)
		Financial markets	(International Monetary Fund, 2020a)
		Chinn - Ito financial openness index	(Chinn and Ito, 2008)
		Ownership of banks,	Fraser Institute (2020b)
		Interest rate controls/negative real interest rates)	Fraser Institute (2020b)
		Credit market regulations	Fraser Institute (2020b)
	Financial institutions depth	Financial institutions depth	(International Monetary Fund, 2020a)
		Bank deposits to GDP (%)	(The World Bank, 2020d)
		Domestic credit to private sector (% of GDP)	(The World Bank, 2020b)
		Private sector credit	Fraser Institute (2020b)
	Financial institutions access	Financial institutions access	(International Monetary Fund, 2020a)
	Financial institutions efficiency	Financial institutions efficiency	(International Monetary Fund, 2020a)
		Bank net interest margin (%)	(The World Bank, 2020d)

		Bank cost to income ratio (%)	(The World Bank, 2020d)
		Bank return on assets (%, before tax)	(The World Bank, 2020d)
		Bank return on equity (%, before tax)	(The World Bank, 2020d)
		Bank concentration (%)	(The World Bank, 2020d)
	Financial institutions stability	Bank Z-score	(The World Bank, 2020d)
		Bank credit to bank deposits (%)	(The World Bank, 2020d)
		Banking crisis dummy	(The World Bank, 2020d)
	Financial markets depth	Financial markets depth	(International Monetary Fund, 2020b)
		Stock market capitalization to GDP (%)	(The World Bank, 2020d)
		Stock market total value traded to GDP (%)	(The World Bank, 2020d)
	Financial markets access	Financial markets access	(International Monetary Fund, 2020a)
	Financial markets efficiency	Financial markets efficiency	(International Monetary Fund, 2020a)
		Stock market turnover ratio (%)	(The World Bank, 2020d)
		Stock market return (%, year on year)	(The World Bank, 2020d)
	Financial markets stability	Stock price volatility	(The World Bank, 2020d)
General economic/ control variables	***General***	Adjusted savings: net national savings (% of GNI)	(The World Bank, 2020f)
		GDP growth (annual %)	(The World Bank, 2020f)
		GDP per capita (current US)	(The World Bank, 2020f)
		Gross fixed capital formation (% of GDP)	(The World Bank, 2020f)
		Industry (including construction), value added (% of GDP)	(The World Bank, 2020f)
		Inflation, consumer prices (annual %)	(The World Bank, 2020f)
		Labour force participation rate, total (% of total population ages 15+) (modelled ILO estimate)	(The World Bank, 2020e)
		Population growth (annual %)	(The World Bank, 2020f)
		Price level of the capital stock, price level of USA in 2011=1	(Feenstra et al., 2015)
		Agriculture, forestry, and fishing, value added (% of GDP)	(The World Bank, 2020f)

		Total natural resources rents (% of GDP)	(The World Bank, 2020f)
		Adults (in 1000)	(Shorrocks et al., 2010, 2011, 2012, 2013, 2014, 2015, 2016. 2017, 2018)
		Median Wealth per Adult (USD)	(Shorrocks et al., 2010, 2011, 2012, 2013, 2014, 2015, 2016. 2017, 2018)
	Technological progress:	ICT goods exports (% of total goods exports),	(The World Bank, 2020f)
		Individuals using the Internet (% of population)	(The World Bank, 2020f)
		Mobile cellular subscriptions (per 100 people)	(The World Bank, 2020f)
		Research and development expenditure (% of GDP)	(The World Bank, 2020f)
		Statistical Capacity score (Overall average)	(The World Bank, 2020f)
Political/ institutional variables	***General***	Civil liberties rating	(Freedom House, 2020)
		Political rights rating	(Freedom House, 2020)
		Status	(Freedom House, 2020)
		Ruling party orientation with respect to economic policy[29]	(Scartascini et al., 2018)
		Ruling party has religious orientation[30]	(Scartascini et al., 2018)
		Control of corruption: estimate	(The World Bank, 2020a)
		Government effectiveness: estimate	(The World Bank, 2020g)
		Political stability and absence of violence/terrorism: estimate	(The World Bank, 2020g)
		Regulatory quality: estimate	(The World Bank, 2020g)
		Rule of law: estimate	(The World Bank, 2020g)
		Voice and accountability: estimate	(The World Bank, 2020g)
		Legal system & property rights	Fraser Institute (2020b)
		Sound money	Fraser Institute (2020b)

[29] I have modified it so that it is a dummy variable with the levels: 1: ruling party has left orientation in terms of economic policy, 0: ruling party has no left orientation.
[30] I have modified it so that it is a dummy variable with levels 1: ruling party has religious orientation, irrespective of which religion, 0: ruling party has no religious orientation.

Inequality/ taxes /redistribution	***General***	General government final consumption expenditure (% of GDP)	(The World Bank, 2020f)
		Gini, income disposable	(Solt, 2020)
		Relative redistribution	(Solt, 2020)
		Transfers and subsidies	Fraser Institute (2020b)
		Government investment	Fraser Institute (2020b)
		Top marginal tax rate	Fraser Institute (2020b)
		State ownership of assets	Fraser Institute (2020b)
		Size of government	Fraser Institute (2020b)
		Labour market regulations	Fraser Institute (2020b)
War/ coup	***General***	Coup[31]	(Powell and Thyne, 2011)
		Military expenditure (% of GDP)	(The World Bank, 2020f)
		War (simple)[32]	(Gleditsch et al., 2002; Pettersson and Öberg)
		War (intense)	(Gleditsch et al., 2002; Pettersson and Öberg)
		War (primary)	(Gleditsch et al., 2002; Pettersson and Öberg)
Education	***General***	Adjusted savings: education expenditure (% of GNI)	(The World Bank, 2020f)
		Education index	(United Nations, 2020)
		Human capital index	(Feenstra et al., 2015)
Entrepreneurship	***General***	Cost of business start-up procedures (% of GNI per capita)	(The World Bank, 2020f)
		Economic freedom summary index	Fraser Institute (2020b)
		Business regulations	Fraser Institute (2020b)
		Regulation	Fraser Institute (2020b)

[31] Coup: I have modified the data so that it is a dummy variable with two levels: 0: no coup attempt in this year and country, 1: coup attempt (regardless of how many or successful or not).

[32] War: I modified each data series so that they are dummy variables: 0: no war (in this form), 1: war.
War (simple): War if, according to the UCDP/PRIO Armed Conflict Dataset, a country participated in a war in that year, irrespective of whether it was a primary or secondary participant as defined in the dataset, and irrespective of whether the conflict had a low or high intensity.
War (primary): only war if country is primary participant, regardless of whether low or high intensity.
War (intense): only war if the conflict in which the country participated was of high intensity, no matter whether it was a primary or secondary participant.

Table A3: Full results of homoscedastic lasso for fixed effects

Selected Variable (in order of selection)	Lasso coefficient value	Post lasso fixed effects estimated coefficient value
Exports of goods and services (% of GDP)	0.0128732	0.0476976
Labour force participation rate, total (% of total population ages 15+)	-0.0234330	-0.3098289
Financial market depth index	0.0222355	0.0314907
Education index of the UN	0.1064835	0.0739295
Domestic credit to private sector (% of GDP)	0.0006787	0.0115851
Human capital index, based on years of schooling and returns to education	0.8411032	2.3817294
Globalisation index economic dimension	0.0220203	0.0729364
Business regulations index	0.0207417	0.0589069

Table A4: Full results of heteroscedastic lasso for fixed effects

Selected Variable (in order of selection)	**Lasso coefficient value**	**Post lasso fixed effects estimated coefficient value**
Exports of goods and services (% of GDP)	0.0126271	0.0548649
Labour force participation rate, total (% of total population ages 15+)	-0.0481800	-0.2771236
Financial market depth index	0.0136009	0.0453204
Education index of the UN	0.1063108	0.1478012
Human capital index, based on years of schooling and returns to education	1.4109806	2.4239423
Control of corruption: estimate	-0.0046511	-0.0945218

Table A5: Regression results multiple robustness checks

	Dependent variable:					
	Wealth Gini				Wealth Shares	
	Countries with direct data	172 countries	3-year averages	6-year averages	Top 1% wealth share	Top 10% wealth share
	(1)	(2)	(3)	(4)	(5)	(6)
Exports/GDP	0.026	0.033**	0.044**	0.048**	0.044*	0.047**
	(0.041)	(0.014)	(0.019)	(0.021)	(0.025)	(0.022)
Labour force participation	-0.320**	-0.223***	-0.254***	-0.234***	-0.365***	-0.313***
	(0.146)	(0.056)	(0.058)	(0.064)	(0.094)	(0.076)
Financial market depth	0.025	0.037**	0.028	0.034	0.038	0.036
	(0.029)	(0.017)	(0.022)	(0.028)	(0.025)	(0.022)
Domestic credit	-0.017	0.002	0.003	0.001	0.005	0.004
	(0.015)	(0.006)	(0.008)	(0.007)	(0.011)	(0.009)
Education	0.254***	0.078***	0.043	-0.009	0.112***	0.106***
	(0.078)	(0.025)	(0.027)	(0.033)	(0.039)	(0.034)
Business regulations	0.162***	0.060***	0.077***	0.122***	0.092***	0.093***
	(0.036)	(0.019)	(0.029)	(0.038)	(0.030)	(0.026)
Control of corruption	-0.168**	-0.065**	-0.095***	-0.132***	-0.072*	-0.069*
	(0.066)	(0.030)	(0.036)	(0.045)	(0.042)	(0.038)
Constant	71.161***					
	(13.556)					
Observations	372	1,862	631	335	1,698	1,698
Within R^2	0.435	0.141	0.197	0.257	0.128	0.146
F Statistic		5.3*** (df = 7; 130)	5.95*** (df = 7; 116)	5.77*** (df = 7; 117)	7.38*** (df = 7; 116)	7.21*** (df = 7; 116)

Note: *p<0.1,**p<0.05,***p<0.01

Table A6: Regression results alternative variables for exports and education

	Dependent variable:
	Wealth Gini
Globalisation economic dimension	0.099*** (0.030)
Labour force participation	-0.225*** (0.065)
Financial market depth	0.033* (0.019)
Domestic credit	0.002 (0.007)
Human capital	2.152** (0.850)
Business regulations	0.070*** (0.025)
Control of corruption	-0.055 (0.036)
Observations	1,631
Within R^2	0.153
F Statistic	5.91*** (df = 7; 1514)
Note:	*p<0.1,**p<0.05,***p<0.01

Table A7: Regression results alternative variables for all variables

	Dependent variable:
	Wealth Gini
Globalisation general[33]	0.119***
	(0.038)
Gini, income disposable	0.211**
	(0.090)
Financial market depth[34]	0.025
	(0.026)
Financial institution depth	2.551
	(4.047)
Education expenditure[35]	0.212**
	(0.181)
Economic freedom summary index	0.749*
	(0.416)
Rule of law: estimate	-1.187
	(0.724)
Observations	1,702
Within R^2	0.083
F Statistic	3.90*** (df = 7; 119)
Note:	*p<0.1,**p<0.05,***p<0.01

[33] This represents yet another measure of globalisation than the one in the previous regression in Table A6. Here it is the aggregate globalisation index of ETH Zurich, whereas in Table A6 it is a sub-index of the globalisation index, namely the globalisation index of the economic dimension. For details, please consult the list of all variables in Table A2.

[34] For alternative variables of financial market depth see Table A8 on the next page.

[35] Adjusted savings: education expenditure (% of GNI)

Table A8: Regression results alternative variables for financial market depth in developed countries sample

	Dependent variable:	
	Wealth Gini	
		(2)
Exports/GDP	0.012	0.023
	(0.026)	(0.032)
Labour force participation	-0.513***	-0.510***
	(0.157)	(0.152)
Domestic credit	-0.004	-0.001
	(0.013)	(0.012)
Education	0.132	0.110
	(0.080)	(0.075)
Business regulations	0.097***	0.099***
	(0.034)	(0.031)
Control of corruption	-0.183**	-0.200**
	(0.073)	(0.077)
Stock market total value traded	-0.003 (0.002)	
Stock market capitalization		-0.004 (0.003)
Observations	437	443
Within R^2	0.255	0.267
F Statistic	5.19*** (df = 7; 30)	5.13*** (df = 7; 30)
Note:		*p<0.1,**p<0.05,***p<0.01

Table A9: Regression results with more control variables

	Dependent variable:	
	Wealth Gini	
	More Control variables 1	More Control Variables 2
	(1)	(2)
Exports/GDP	0.032*	0.034*
	(0.019)	(0.019)
Labour force participation	-0.201***	-0.173**
	(0.075)	(0.078)
Financial market depth	0.040**	0.032*
	(0.019)	(0.017)
Domestic credit	0.006	0.005
	(0.009)	(0.009)
Education	0.122***	0.089*
	(0.033)	(0.047)
Business regulations	0.077***	0.066***
	(0.022)	(0.020)
Control of corruption	-0.079**	-0.082**
	(0.035)	(0.035)
GDP per capita (current US)	-0.00004*	-0.0001**
	(0.00002)	(0.00002)
Agriculture[36]	0.008	0.030
	(0.062)	(0.063)
Gini, income disposable	0.179**	0.181**
	(0.081)	(0.092)
Adults (in 1000)		0.00002**
		(0.00001)
Inflation		-0.027**
		(0.014)
Internet users[37]		0.010
		(0.014)
Observations	1,508	1,454
Within R^2	0.164	0.184
F Statistic	4.32*** (df = 10; 113)	6.19*** (df = 13; 111)
Note:		*p<0.1,**p<0.05,***p<0.01

[36] Agriculture, forestry, and fishing, value added (% of GDP)

[37] Individuals using the Internet (% of population)

Table A10: Regression results two-way fixed effects baseline regression

	Dependent variable:
	Wealth Gini
Exports/GDP	0.034*
	(0.019)
Labour force participation	-0.234***
	(0.057)
Financial market depth	0.033*
	(0.019)
Domestic credit	0.004
	(0.007)
Education	0.026
	(0.052)
Business regulations	0.058**
	(0.027)
Control of corruption	-0.067**
	(0.031)
Dummy year 2002	-0.494*
	(0.291)
Dummy year 2003	-0.142
	(0.322)
Dummy year 2004	0.443
	(0.469)
Dummy year 2005	0.255
	(0.512)
Dummy year 2006	0.226
	(0.524)
Dummy year 2007	0.442
	(0.547)
Dummy year 2008	-0.296
	(0.587)
Dummy year 2009	0.004
	(0.557)
Dummy year 2010	-0.299
	(0.573)
Dummy year 2011	-0.101
	(0.634)

Dummy year 2012	0.613
	(0.657)
Dummy year 2013	1.358*
	(0.720)
Dummy year 2014	0.565
	(0.743)
Dummy year 2015	0.737*
	(0.763)
Dummy year 2016	0.759*
	(0.783)
Dummy year 2017	-0.006
	(0.829)
Observations	1,698
Within R^2	0.186
F Statistic	9.02*** (df = 23; 116)
Note:	*p<0.1,**p<0.05,***p<0.01

Table A11: Regression results baseline without education or with education expenditure

	Dependent variable:	
	Wealth Gini	
	No education	Education expenditures
	(1)	(2)
Exports/GDP	0.041^{**}	0.041^{**}
	(0.017)	(0.017)
Labour force participation	-0.240^{***}	-0.239^{***}
	(0.058)	(0.058)
Financial market depth	0.045^{**}	0.045^{**}
	(0.017)	(0.018)
Domestic credit	0.006	0.005
	(0.008)	(0.008)
Business regulations	0.087^{***}	0.087^{***}
	(0.022)	(0.022)
Control of corruption	-0.077^{**}	-0.076^{**}
	(0.032)	(0.032)
Education expenditures		0.126
		(0.173)
Observations	1,700	1,671
Within R^2	0.140	0.140
F Statistic	6.15^{***} (df = 6; 116)	5.21^{***} (df = 7; 114)
Note:		*p<0.1,**p<0.05,***p<0.01

Table A12: Regression results omitted independent variables

	Dependent variable:					
	Wealth Gini					
	No exports	No labour force participation	No financial market depth	No Domestic credit	No business regulations	No control of corruption
	(1)	(2)	(3)	(4)	(5)	(6)
Exports/GDP		0.039^{**}	0.045^{***}	0.043^{***}	0.045^{***}	0.039^{**}
		(0.017)	(0.017)	(0.016)	(0.017)	(0.017)
Labour force participation	-0.229^{***} (0.057)		-0.239^{***} (0.062)	-0.215^{***} (0.051)	-0.251^{***} (0.066)	-0.250^{***} (0.059)
Financial market depth	0.053^{***} (0.017)	0.046^{**} (0.019)		0.039^{**} (0.018)	0.049^{**} (0.022)	0.045^{**} (0.018)
Domestic credit	0.006 (0.007)	0.003 (0.008)	0.010 (0.008)		0.010 (0.008)	0.007 (0.008)
Business regulations	0.091^{***} (0.023)	0.090^{***} (0.024)	0.088^{***} (0.023)	0.086^{***} (0.020)		0.077^{***} (0.022)
Control of corruption	-0.075^{**} (0.032)	-0.074^{**} (0.034)	-0.080^{**} (0.032)	-0.069^{**} (0.029)	-0.043 (0.032)	
Observations	1,708	1,700	1,708	2,051	1,796	1,782
Within R^2	0.120	0.110	0.124	0.133	0.091	0.129
F Statistic	7.05^{***} (df = 5; 117)	5.44^{***} (df = 5; 116)	6.37^{***} (df = 5; 118)	7.9^{***} (df = 5; 125)	5.19^{***} (df = 5; 118)	6.49^{***} (df = 5; 116)
Note:						*p<0.1,**p<0.05,***p<0.01